ERGONOMICS, 15 JULY, 2004, VOL. 47, NO. 9, 965–986

Auditory alerts for in-vehicle information systems: The effects of temporal conflict and sound parameters on driver attitudes and performance

EMILY E. WIESE and JOHN D. LEE*

Department of Mechanical and Industrial Engineering, The University of Iowa, Iowa City, IA, USA

Keywords: Telematics; Warnings; Alerts; Urgency; Annoyance; Distraction

In-vehicle information systems will soon confront drivers with an increasing number of warnings and alerts for situations ranging from imminent collisions to the arrival of e-mail messages. Coordinating these alerts can ensure that they enhance rather than degrade driving safety. Two experiments examined how temporal conflict and sound parameters affect driver performance and acceptance. The temporal conflict of an e-mail alert occurring 300 ms before a collision warning interfered with the response to the collision warning, but an e-mail alert occurring 1000 ms before the collision warning had the opposite effect and enhanced the response to the collision warning. These results emphasize the need to consider how in-vehicle devices influence drivers' strategic anticipation of high-demand situations. Regarding sound parameters, results showed that highly urgent sounds tended to speed drivers' accelerator release, but the annoyance associated with highly urgent sounds increased workload. In fact, there was a strong positive association between ratings of annoyance and subjective workload. Consistent with the urgency mapping principle, there was a slight negative association between the differences in the rated urgency of collision warnings and e-mail alerts and subjective workload. The results suggest that warning and alert design should consider an annoyance trade-off in addition to urgency mapping.

1. Introduction

As in-vehicle information systems proliferate, an increasingly important design challenge is to coordinate the many different auditory and visual alerts being presented to the driver to ensure that they enhance rather than degrade driving safety. Auditory warnings are a particularly appealing design option because they may distract drivers less than visual displays. As a result, in-vehicle information systems may include a variety of sounds to alert drivers to situations ranging from an imminent collision to the arrival of a new e-mail message. However, little research has addressed how a lack of system coordination might cause these sounds to compete for a driver's attention. A relatively unimportant sound, such as an e-mail alert, given in close temporal proximity to a relatively important sound, such as a collision avoidance warning, might cause confusion and undermine driving safety. Current in-vehicle information system design tends to consider each system individually, and does not consider how to manage the combined demand on the

*Author for correspondence. e-mail: jdlee@engineering.uiowa.edu

Ergonomics ISSN 0014-0139 print/ISSN 1366-5847 online
http://www.tandf.co.uk/journals
DOI: 10.1080/00140130410001686294

driver. A design approach that considers the integration of these systems may be critical.

An important aspect of integrating in-vehicle information systems is the timing of various alerts and warnings. Concurrent performance of two tasks (such as responding to e-mail messages and driving) often results in poorer performance of either or both tasks, depending on the levels and types of demands of each task and their priorities (Wickens 1984, 2002, Haigney and Westerman 2001). Mental workload and the dual task interference associated with the concurrent performance of tasks that share the same mental resources, such as listening to two auditory alerts, tends to be higher than performing tasks that do not compete for the same resources. If the tasks share the same perceptual mode and response mode then resource competition is even greater. The psychological refractory period (PRP) effect, a form of dual-task interference, refers to the delayed response a person has to a second stimuli when the period between it and an initial stimuli is very close in time (Pashler 1994). The bottleneck theory of Pashler suggests that the information processing stage of response selection limits dual-task performance by acting as a bottleneck. Dual-task studies have shown that a stimulus onset asynchrony of 300 ms or less interferes maximally with the subsequent stimulus response generation (Pashler 1994). Coordinating the onset of alerts to avoid temporal conflict may be a critical consideration in integrating in-vehicle information systems.

The specific characteristics of auditory alerts and warnings might extend the signal interpretation and response selection processes and exacerbate the delay associated with the PRP effect or resource competition. Alerts whose meanings are ambiguous could delay response and increase subjective workload. Like the automotive environment, the medical environment faces a proliferation of auditory warnings and the ambiguous meaning of the alerts has undermined performance and increased workload (Meredith and Edworthy 1995). In response to this problem, Edworthy (1994) introduced the urgency mapping principle, which states that the urgency of the situation should match the perceived urgency of the alert. Urgency of sound may guide attention and help people give a higher priority to highly urgent events when responding. For driving situations, appropriate urgency mapping pairs a low-urgency sound to an e-mail system and a high-urgency sound to a collision avoidance system. This should enhance warning interpretation, reduce interference, and result in lower workload. An inappropriate urgency mapping would do the reverse—pairing a high-urgency sound to the e-mail system and a low-urgency sound to the collision avoidance system would lead to more ambiguity and higher workload. An inappropriate urgency mapping might undermine the response to the collision warning. Specifically, an urgent sounding e-mail alert immediately followed by a less urgent collision avoidance warning might delay a braking response. Mapping highly urgent sounds to collision avoidance warnings and low-urgency sounds to e-mail alerts could reduce the confusion that might arise from temporal conflict. Previous studies have shown that certain sound parameters systematically affect rated urgency (Edworthy *et al.* 1991, 1995), but none have examined how it might reduce the demands of interpreting multiple warnings in a realistic driving environment. Beyond the effect on driving performance, inappropriately mapped alerts could also annoy drivers and undermine their acceptance of in-vehicle information systems. Urgency mapping represents a potentially useful way to integrate alerts and warnings in the automotive environment.

Two experiments examined the effects of auditory alerts on driver attitudes and performance. Both use auditory stimuli that required a response. The auditory stimuli were a rear-end collision avoidance warning, which required a braking response, and an e-mail system alert, which required the driver to respond vocally. According to the PRP and multiple resource theories, an e-mail alert issued immediately before a collision avoidance warning might delay drivers' braking response. Appropriate urgency mapping may mitigate this effect, while inappropriate mapping may exacerbate it. The first experiment examined how sound parameters and temporal conflict affect rated urgency, annoyance, appropriateness, and workload, as well as drivers' braking performance. The second study expanded the range of temporal conflict and sound parameters to clarify the results of the first experiment. Both experiments addressed how sounds from two distinct systems interact to affect driver response.

2. Experiment 1

The first experiment tested several specific hypotheses: (1) sound parameters will affect rated urgency, annoyance, and appropriateness; (2) inappropriate urgency mapping will increase workload (that is, the pairing of an urgent e-mail alert with a less urgent collision warning will increase subjective workload); (3) highly urgent warnings will reduce drivers' reaction time to a braking lead vehicle; and (4) the e-mail alert will interfere with the collision avoidance response, and the high-urgency e-mail alert will be more disruptive than the low-urgency alert.

2.1. *Method*

2.1.1. *Participants*: Sixteen drivers between the ages of 18 and 35 participated in this study. All participants were screened to eliminate those prone to simulator sicknesses. Exclusion criteria included possible pregnancy and experience with migraine headaches, claustrophobia, or motion sickness. Participants were compensated a maximum amount of $45.00. They received $15.00 upon completion of the first day of the experiment and $25.00 upon completion of the second day. To encourage drivers to take the e-mail task seriously, they received a bonus of $5.00 at the end of the two day experiment if all the tasks were completed appropriately. Information regarding the participants' prior experience with computer technology, and e-mail in particular, was not gathered. However, no participants expressed any concern about the task or experienced difficulties performing the e-mail task.

2.1.2. *Apparatus*: A fixed-based, medium-fidelity driving simulator was used to conduct this study. The driving simulator uses a 1992 Mercury Sable configured with the Vection Research Simulator (VRS). The VRS is a fully integrated, high-performance driving simulation designed for use in ground vehicle research and training applications. The VRS uses a real Mercury Sable vehicle cab that has been modified to include a 50-degree visual field of view, full instrumentation with actual gauges, force feedback steering wheel, and a rich audio environment. Fully textured graphics are generated by state-of-the-art PC hardware that delivers a 60-Hz frame rate at 1024×768 resolution. All graphics for roadway layouts, markings, and signage conform to American Association of State Highway and Transportation Officials (AASHTO) and Manual of Uniform Traffic Control Devices (MUTCD) design standards. The ambient sound level of the simulator during the experiment was 54 dBA.

2.1.3. *Independent and dependent variables*: The context and characteristics of the sounds used to alert drivers to an imminent collision or to the receipt of e-mail defined the independent variables. The context was defined by the accuracy of the collision avoidance system warning (hit or false alarm) and by the e-mail alert timing (300 ms before a collision avoidance warning [referred to as 'present'], or non-conflicting [referred to as 'absent']). The false alarms comprised 50% of the collision warnings and were randomly distributed throughout each drive. The sound characteristics included the sonic urgency of the collision avoidance warning (defined by varying the burst density and loudness) and of the e-mail alert (defined by varying the onset and offset rates). Sonic urgency refers to the urgency of the sound based on the sound parameters that previous research has shown to influence perceived urgency as measured by subjective ratings or reaction time to warnings (Edworthy *et al.* 1991, Hellier *et al.* 1993, Edworthy and Adams 1996, Marshall *et al.* 2001). In this paper, we refer to sonic urgency as 'urgency' and perceived urgency as referred to as 'rated urgency'. The collision warning was developed and tested for collision warning systems (Lerner 1991, Kiefer *et al.* 1999). The e-mail alert was developed and tested as part of a previous set of experiments (Marshall *et al.* 2001) The sound parameters that we chose to define high and low levels of urgency were those that have shown a strong effect on perceived urgency in previous research (Edworthy and Adams 1996, Marshall *et al.* 2001).

Each pair of sounds was constructed to produce a clear difference in sonic urgency. A series of studies identified several variables that affect the perceived urgency of warnings as measured by subjective ratings (Marshall *et al.* 2001). These studies provided participants with a driving scenario for which the tones were presented. The participants' task was to rate the urgency and annoyance of warnings with respect to a specific scenario. The results showed that manipulating burst density onset and offset rates had a powerful influence on the rated urgency of sounds. Accordingly, the collision avoidance warnings developed for this experiment differed according to burst density and loudness. Loudness was also used to define the sounds because previous research has demonstrated it to have a large and consistent effect on the rated urgency and annoyance of signals (Edworthy and Adams 1996). Thus, loudness was combined with burst density to create two collision warnings with differing sonic urgency that matched the relatively high nominal urgency of the collision avoidance situation. The low-urgency collision warning was composed of three pulses of sound of 0.107 s duration, interspersed with 0.007 s of silence. At the end of the third pulse, there was a 0.113 s period of silence before the next burst began. The burst density for the low-urgency collision avoidance warning was 0.76. The density for the high-urgency warning was 0.94. This warning featured a continuous sound pulse of 0.107 s duration interspersed with 0.007 s of silence. There were 20 consecutive bursts in the sound. This warning was 17 dB louder than the low-urgency warning, which was set to 61 dBA. Although guidelines regarding auditory warnings suggest a loudness of at least 15 dB above the ambient noise level, a lower level was chosen for this experiment. A lower level was acceptable for this experiment because the frequency spectrum of the warning was well separated from the ambient noise and several other experiments demonstrated that the 61 dBA warning could be heard and that it effectively alerted drivers and helped them avoid collision situations (Lee *et al.* 2002).

The e-mail alert was varied according to onset and offset of the sound, a parameter that was also found to greatly affect both urgency and annoyance (Marshall *et al.*

2001). Onset and offset were used for e-mail alerts, rather than pulse density or loudness, so that comparisons between e-mail alerts and collision warnings would reflect the underlying construct of urgency rather than surface features of the auditory signal. Additionally, varying onset and offset for the e-mail alert helped match sonic urgency to the nominal urgency of the arrival of a new e-mail message. The e-mail alert had a low burst density (0.2) and was based on the high /i/ frequency series. The low-urgency warning featured a slow onset and offset, with the sound fading in during the first half of the warning and fading out during the last half of the warning. The high-urgency e-mail alert had no onset or offset (or fade-in and fade-out). The designation of high and low urgency refers to the assumed high and low sonic urgency produced by the sound parameters. All independent variables were within-subject variables, and all drivers experienced all combinations of the independent variables. Table 1 describes the experimental conditions.

The dependent variables included subjective measures of urgency, annoyance, appropriateness, and workload. Drivers provided subjective ratings in a post-drive questionnaire in which they rated the urgency, annoyance, and appropriateness of both the collision avoidance and e-mail alerts on a scale of 0 to 10. Subjective mental workload was measured using the NASA TLX workload scale (Hart and Staveland 1988).

Other dependent variables characterized braking performance. The decomposed reaction time was of particular interest. This variable was composed of three measures: accelerator release reaction time, accelerator to brake transition time, and brake to maximum brake transition time. The accelerator release reaction time is the reaction time to the onset of the lead vehicle braking, measured from the moment the lead vehicle applies the brakes, which coincided with the illumination of the brake lights of the lead vehicle. The accelerator to brake transition time is the time from accelerator release to brake application. The final component, brake to maximum brake transition time, is the time from initial brake application to maximum deceleration. The mean deceleration and the maximum deceleration during braking describe the braking profile. The mean deceleration is the average deceleration of the vehicle from initial brake press until the braking event ends or the driver's vehicle collides with the lead vehicle. The maximum deceleration is defined as the peak deceleration between the start and end of the braking event. Safety benefit measures that quantify the effect of the warnings on collision and collision severity were also examined. Collisions refer to the number of collisions that occurred. The collision severity is measured by the velocity at the time of impact. However, as in actual driving situations, collisions were rare and only seven collisions occurred. No safety

Table 1. Summary of experimental conditions

Independent variable	Conditions
Collision warning accuracy (within)	Accurate identification of a collision situation (Hit) False alarm (FA)
E-mail alert timing (within)	300 ms before the collision warning (Present) Not concurrent with collision warning (Absent)
Collision warning urgency (within)	High burst density (0.94), Loudness = 78 dBA Low burst density (0.76), Loudness = 61 dBA
E-mail alert urgency (within)	No onset or offset Slow onset and slow offset (50% of warning combined)

benefit measures showed a statistically significant effect for the first experiment and are not discussed.

2.1.4. *Experimental design and protocol*: Each participant drove in the simulator for 1.5 h a day for 2 days, not necessarily at the same time each day. During this time, the participants drove eight scenarios that depicted a straight, level, rural two-lane highway. A moderate amount of ambient traffic in the opposing lane discouraged drivers from passing the lead vehicle during the braking events. Each participant drove one 6 min practice scenario and four 12 min scenarios each day. The eight experimental scenarios were all 18.4 km long, while the practice scenario was 7.0 km in length. The practice scenario allowed the driver to adapt to the simulator dynamics and was administered each day before the experimental scenarios. The collision warning and e-mail alert were each presented to the driver twice during the 6 min practice drive for a total of four occurrences that enabled the driver to learn the meaning of each sound.

In each scenario, drivers received two collision avoidance warnings and four e-mail alerts. One of the collision avoidance warnings was associated with a braking lead vehicle (hit) and the other a false alarm (FA). One of the e-mail alerts sounded 300 ms before one of the collision avoidance warnings (present), and the other three sounded independently of the collision avoidance warnings (absent). When the e-mail alert was ‘present’, it was paired with either a hit or false alarm collision avoidance warning. All combinations of the low- and high-urgency collision warning and e-mail alerts and present and absent e-mail alerts created a total of eight scenarios. Each driver experienced a total of 16 collision warnings and 32 e-mail alerts. Table 2 shows the experimental conditions represented in each scenario. The warning algorithm used to trigger the collision avoidance warning during the braking events was developed by Burgett *et al*. (1998), and generates a warning based on the kinematics of the collision situation. The precise timing depended on the relative speed and distance of the two vehicles, but generally occurred within 300 ms after the lead vehicle began to brake. The scenario order was counterbalanced using

Table 2. Experimental conditions represented in the eight scenarios

Scenario	CW condition	E-mail condition	CW warning urgency	E-mail alert urgency
1.1	FA	Present	High	High
	Hit	Absent		
1.2	Hit	Present	High	High
	FA	Absent		
1.3	FA	Present	High	Low
	Hit	Absent		
1.4	Hit	Present	High	Low
	FA	Absent		
2.1	Hit	Present	Low	High
	FA	Absent		
2.2	FA	Present	Low	High
	Hit	Absent		
2.3	FA	Present	Low	Low
	Hit	Absent		
2.4	Hit	Present	Low	Low
	FA	Absent		

a digram-balanced 4 × 4 Latin square. The resulting combination of day 1 and day 2 scenario orders resulted in 16 different conditions. Each participant was randomly assigned to a condition. The location of the collision avoidance warnings and e-mail alerts were randomized within each of the eight scenarios.

The e-mail task required drivers to respond to certain messages and delete others, depending on the message content. Before each scenario, drivers were given instructions and a description of the task they were to perform. The drivers were told that the car was equipped with an advanced information system that included a voice-activated e-mail system and a collision warning system. They were also told that they had recently placed bids on a CD player, computer monitor, and a mountain bike on eBay™, an online auction company, and that the auction was in its final stages. The drivers were told that eBay™ had sent an e-mail to notify them if they had been outbid. If they had been outbid, they were to respond to the e-mail and automatically increment the bid. The drivers were told they had placed bids on several other items as well; however, they were instructed to increment the bids only on the CD player, computer monitor, and the mountain bike. Regarding the collision warning system, drivers were told that the collision warning system detects situations when the lead vehicle brakes suddenly, but that it is imperfect and periodically issues false warnings.

The drivers were required to memorize only one command, 'Auto E-mail On', which turned on the system. All other commands were given to the drivers after they had spoken this command. When the e-mail alert indicated that a new message had been received, the drivers turned on the system by saying 'Auto E-mail On'. The e-mail system then responded by listing four options available to the driver:

- 'Read Messages': the command to read any new messages received.
- 'Autobid': allowed the driver to automatically respond to the auction notification with a higher bid.
- 'Delete': deleted the current message.
- 'Auto E-mail Off': turned the e-mail system off.

After each scenario, drivers completed subjective workload ratings and a post-drive questionnaire. After the experiment was complete, the experimenter debriefed the drivers and solicited general comments.

2.2. *Results*

2.2.1. *Rated urgency*: The sound parameters for the collision warning and e-mail alert failed to influence rated urgency. Varying burst density and loudness did not result in significant differences in the mean urgency ratings of the collision warnings, $F(1,15) = 0.40$, $p = 0.535$. Surprisingly, the high-urgency collision warning was rated as 4.93 and the low-urgency collision warning was rated as 4.98. There were also no significant effects for the urgency ratings of the e-mail alerts, $F(1,15) = 0.45$, $p = 0.5140$. Mean ratings for the high- and low-urgency e-mail alerts were 4.19 and 4.07, respectively. As was expected, the collision warnings were rated as more urgent (4.98) than the e-mail alerts (4.13), $F(1,226) = 11.53$, $p = 0.0003$.

2.2.2. *Rated annoyance*: The burst density and loudness affected annoyance of the collision avoidance warning, $F(1,15) = 13.15$, $p = 0.0025$. As was expected, the high-urgency collision avoidance warning was more annoying than the low-urgency warning, with a mean rating of 5.30 compared to 4.35. However, a high- and low-urgency e-mail alert did not have a significant effect on the perceived annoyance of the e-mail alert, $F(1,15) = 0.54$, $p = 0.4729$. Overall, the collision warnings were rated as more annoying than the e-mail alerts, $F(1,226) = 39.72$, $p < 0.0001$, with a mean rating of 4.98 for the collision warnings and 3.86 for the e-mail alerts. This is particularly striking in that drivers received twice as many e-mail alerts as collision avoidance warnings.

2.2.3. *Rated appropriateness*: The warning sound parameters of loudness and burst density affected the perceived appropriateness of the collision avoidance warning, $F(1,15) = 7.07$, $p = 0.0179$. The high-urgency warning was perceived as less appropriate than the low-urgency warning, with mean ratings of 4.86 and 5.59, respectively. The ratings of the e-mail alert sound parameters of onset and offset did not produce a significant difference in appropriateness, $F(1,15) = 0.41$, $p = 0.5296$. Figure 1 shows the subjective ratings for urgency, annoyance, and appropriateness for the collision warnings and e-mail alerts. The correlation between annoyance and appropriateness for the collision warning alerts was -0.276, $p = 0.0008$. The correlation between urgency and appropriateness for the collision warning alerts was also significant, 0.587, $p < 0.0001$.

2.2.4. *Subjective workload*: The NASA TLX subscales were combined in an equally weighted average. Collision warning parameters significantly affected rated workload, $F(1,15) = 9.65$, $p = 0.0072$. The high-urgency collision warnings induced higher levels of workload (2.08 on a scale of $0-7$) than low-urgency warnings (1.78). In contrast, the high- and low-urgency e-mail alerts had no statistically significant effect, $F(1,15) = 1.59$, $p = 0.2272$. According to the urgency mapping principle, the combination of the high-urgency collision warning and low-urgency e-mail alert should result in the lowest workload. However, the workload ratings for the warning combinations were not significantly different, $F(3,6) = 2.94$, $p = 0.1208$. Similarly, according to the urgency mapping principle, the difference between the rated urgency of the collision avoidance warning and e-mail alert should be negatively correlated with workload—large difference in rated urgency might reduce the ambiguity in the

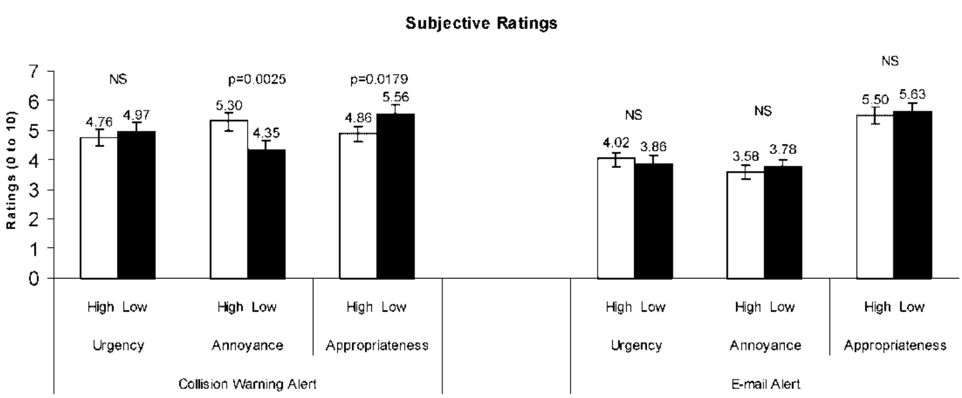

Figure 1. Subjective ratings for the high- and low-urgency collision warnings and e-mail alerts.

sounds and make it easier to identify the appropriate response. The correlation between the difference in rated urgency of the collision warning and the e-mail alert and subjective workload was -0.028, $p > 0.05$. In contrast, the correlation between workload and sum of the rated annoyance for the collision avoidance warning and e-mail alert was 0.227, $p < 0.01$. These results suggest that to reduce workload, urgency mapping may be less important than reducing the annoyance associated with warnings and alerts.

2.2.5. *Driving performance during lead vehicle braking events*: The data from all drivers were combined to form a database containing 128 imminent collision situations. Data were missing from seven cases due to drivers driving too slowly to experience the effect of the lead vehicle deceleration, making reaction times impossible to calculate. Events in which the lead vehicle braked are referred to as braking events.

The accelerator release reaction time for events in which a braking event occurred showed significant effects for the collision warning parameters ($F(1,14) = 12.74$, $p = 0.0031$), the e-mail alert parameters ($F(1,14) = 10.90$, $p = 0.0052$), and the e-mail condition ($F(1,14) = 8.69$, $p = 0.0106$). Figure 2 shows that the high-urgency collision warning produced a mean reaction time that was approximately 325 ms faster than the low-urgency collision warning. The high-urgency e-mail alert also produced a 350 ms faster reaction time to the braking lead vehicle than the low-urgency alert. The presence of the e-mail alert delayed the drivers' response to the lead vehicle braking event by 240 ms. Based on these results, it would be expected to see a significant negative correlation between the rated urgency of the collision warning and the accelerator release reaction time. This was in fact the case, with a correlation of -0.205, $p < 0.05$, even though no significant effects were seen in the collision warning's rated urgency as a function of the warning sound parameters.

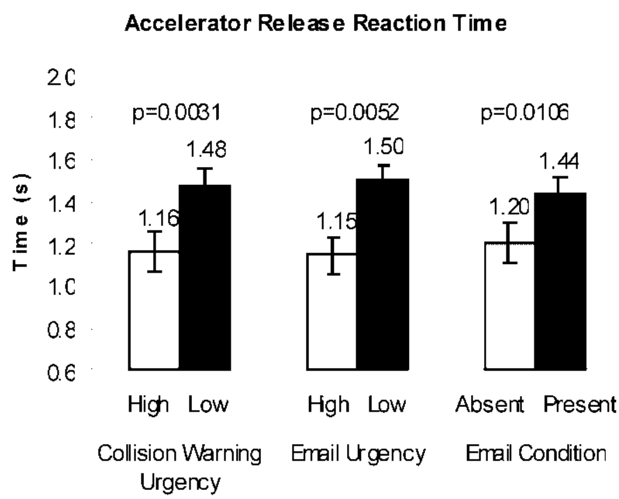

Figure 2. Mean accelerator release reaction times for the collision warning urgency, e-mail alert urgency, and e-mail condition during lead vehicle braking events.

In addition to these main effects, Figure 3 shows that the effects of the e-mail presence and e-mail alert urgency are complicated by an interaction, $F(1,14) = 18.43$, $p < 0.0001$. When the e-mail task did not coincide with the collision warning, the reaction time to the collision warning was substantially faster with the high-urgency e-mail alert; however, when the e-mail coincided with the collision warning the reaction times for the high- and low-urgency e-mail alert were similar.

The collision warning urgency affected the accelerator to brake transition time, $F(1,14) = 11.26$, $p = 0.0047$. The high-urgency collision warning was also associated with a 200 ms slower accelerator to brake transition time than the low-urgency collision warning (780 ms compared to 584 ms). The collision warning urgency also affected drivers' maximum deceleration during lead vehicle braking events, $F(1,14) = 5.35$, $p = 0.0364$. The high-urgency collision warning led to a lower maximum deceleration (0.53 g) than the low-urgency collision warning (0.55 g). This is consistent with the effect on the accelerator release reaction time and accelerator to brake transition time. The high-urgency alert appears to have caused drivers to remove their foot from the accelerator more quickly and then allowed for a more gradual and controlled deceleration. These results are consistent with previous research that shows that an early accelerator release leads to a slower transition time because drivers who released the accelerator early were able to move to the brake in a slower and more controlled fashion (Lee *et al.* 2002). The collision warning urgency, e-mail alert urgency, and e-mail condition showed no significant effect on the brake to maximum brake reaction time.

2.3. *Discussion*

2.3.1. *Rated urgency, annoyance, and appropriateness*:
Although drivers rated collision warnings as more urgent than e-mail alerts, the sound parameters had no significant effects on the urgency ratings of either. One explanation for this result is that the sounds were not sufficiently different. Previous studies investigating sound density found that this parameter dramatically affected the rated urgency of the warning (Marshall *et al.* 2001). The sound densities investigated in those studies ranged from 0.2 to 0.8. This study used a narrower range of densities, 0.76 to 0.94. It

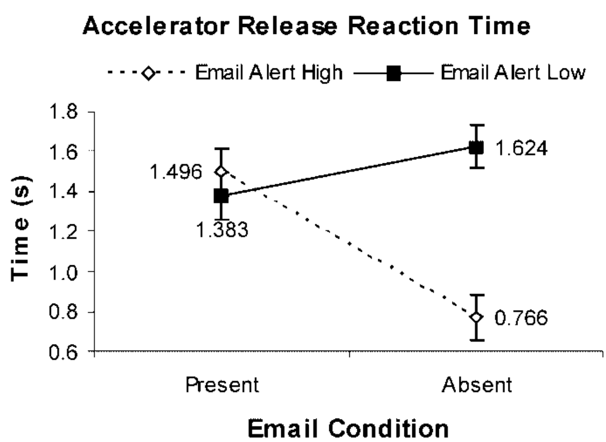

Figure 3. Mean accelerator release reaction times for two-way interaction between the e-mail alert and the e-mail condition. Error bars represent standard errors.

is possible that this range was too small to influence rated urgency. However, this explanation fails to explain the lack of an effect for the e-mail alert. The e-mail alerts differed according to alert onset and offset, which were manipulated to the same degree as in previous experiments (Marshall *et al*. 2001). Another explanation is that rated urgency is a relative rather than absolute judgment and the drivers did not immediately rate the warnings upon hearing them. Urgency may be best rated in relation to other sounds rather than on an absolute basis. Another explanation that has important practical implications for warning design is that the context plays an important role in drivers' interpretation of warnings and alerts. It may be that drivers simply classified e-mail alerts as low-urgency independent of the sound parameters and likewise classified collision warnings as high-urgency independent of the sound parameters.

Unlike rated urgency, rated appropriateness and annoyance did show effects of the sound parameters. This experiment shows a stronger relationship between sound parameters and rated annoyance than between sound parameters and rated urgency. As expected, the high-urgency collision warning was rated more annoying than the low-urgency collision warning, and the collision warnings were rated more annoying than the e-mail alerts. In addition, the burst density in the collision avoidance warning affected subjective workload. The subjective workload did not respond as might have been predicted by Edworthy's urgency mapping principle. According to this principle, the high-urgency collision warning and low-urgency e-mail alert combination should have produced the least workload, and the low-urgency collision warning and high-urgency e-mail alert combination the greatest workload. Likewise, differences between the rated urgency of collision warnings and e-mail alerts are less correlated with workload compared to the ratings of annoyance. Highly annoying warning combinations tended to receive the highest workload rating, suggesting that annoyance rather than urgency mapping may affect workload most strongly. As in previous experiments, these results show that a trade-off exists between urgency and annoyance (Marshall *et al*. 2001). These results suggest that warning design should go beyond urgency mapping and consider perceived annoyance.

2.3.2. *Effects of sound parameters on driving performance*: The sound parameters affected driving performance even though they did not affect the ratings of urgency. The high-urgency collision warning induced a faster accelerator release. Interestingly, the high-urgency collision warning resulted in a slower accelerator to brake transition time than the low-urgency collision warning. Likewise, the maximum deceleration during lead vehicle braking events was lower for the high-urgency warning. The effect of warning urgency on the response process reflects the faster accelerator release that allows the other components of the response to proceed in a slower, more controlled manner that results in a less severe collision situation. A slow accelerator release requires a more abrupt deceleration to avoid a collision. These results are consistent with a previous collision warning study and are particularly interesting because the previous study was conducted in a motion-base simulator that provided substantial vestibular braking cues that were absent in the present study (Lee *et al*. 2002).

2.3.3. *Effects of alert interference*: The e-mail alerts interfered with drivers' response to collision warnings, and a two-way interaction shows that the e-mail alert urgency mediated the effect of the alert on interference. When the e-mail task coincided with the lead vehicle braking event, the accelerator release reaction times for both the high- and low-urgency e-mail alerts were nearly equal. In contrast, when the e-mail task was absent, the accelerator release time for the high-urgency e-mail alert was over 850 ms faster than the accelerator release time for the low-urgency alert. It seems that high-urgency alerts are particularly disruptive compared to low-urgency alerts when they coincide with collision warnings.

2.3.4. *Unresolved issues*: Although the results suggest that sound parameters and alert timing can affect driver attitudes and performance, several issues merit further investigation. The lack of support for the hypothesis that sound parameters affect rated urgency may reflect the relatively small range of burst density used in the experiment. Selecting warnings with more distinct burst densities might lead to greater differences in rated urgency. The main effect of the presence of the e-mail alert on the reaction time to the lead vehicle is consistent with the multiple resource theory and PRP theories of dual task interference; however the main effect of e-mail alert urgency and the interaction between urgency and presence of the alert is difficult to explain—the high-urgency e-mail alert is associated with faster reaction times when the e-mail alert does not coincide with the collision warning. The e-mail response selection is more complex than typical PRP selection tasks. The interference associated with the 300 ms difference between the e-mail alert and the collision warning may reflect the initial processing of the two signals. Interference may be greater if the e-mail alert occurred earlier so that the interference is between the response generation in the e-mail task and the perception of the collision warning. Investigating different stimulus onset timings may clarify these results.

3. Experiment 2

The second experiment addressed several issues left unresolved by the first experiment. In Experiment 1, contrary to the hypothesis and the results of previous experiments (Marshall *et al.* 2001), sound density and loudness failed to affect rated urgency. One possible reason for this is that the collision warnings used in the first experiment were too similar in sound density to produce an effect. Another possibility is that the demands of a realistic driving setting do not allow drivers to make precise judgments of alert urgency. Experiment 2 differentiates between these possibilities by using warnings that generated large and highly significant differences in subjective ratings of urgency in a controlled setting. Experiment 1 also found that the occurrence of a high-urgency e-mail alert 300 ms prior to a collision avoidance warning delayed the braking response, but only relative to the situation in which high-urgency e-mail alerts did not coincide with the braking event. It is possible that a longer interval between the e-mail alert and the collision warning could delay the drivers' braking response more severely. This experiment addresses these issues.

3.1. *Method*

3.1.1. *Participants*: Sixteen drivers between the ages of 18 and 35 participated. None of these drivers participated in the previous experiment. As in the first experiment, all participants were screened to eliminate those prone to simulator sickness. Participants were compensated \$20.00 for their participation. They received \$15.00 upon completion of the experiment and a bonus of \$5.00 if all e-mail tasks were completed appropriately.

3.1.2. *Apparatus*: The driving simulator configuration was exactly the same as in the first experiment.

3.1.3. *Independent and dependent variables*: The independent variables were the same as those in Experiment 1 with a few exceptions. This experiment focused on the collision warning parameters, so e-mail alert urgency was not manipulated. The high-urgency alert from Experiment 1 was used, in which the onset and offset of the alert was immediate rather than gradual. The timing of the e-mail alert was modified so that the alert occurred 1000 ms before the collision warning, rather than 300 ms before as in Experiment 1.

The collision warnings were constructed so as to maximize the difference in urgency by varying the burst density and warning duration. The low-density collision warning had a density of 0.20 and the high-density warning had a density of 0.80. In a previous study, this pairing of warnings produced urgency ratings of 38.31 and 61.44—this difference is both highly statistically and practically significant (Marshall *et al*. 2001). We also manipulated warning duration, which also produced highly statistically and practically significant differences in a previous study (Marshall *et al*. 2001). The short duration collision warning had a length of 0.44 s, whereas the long duration collision avoidance warning had a length of 2.26 s. Table 3 summarizes the experimental conditions used in Experiment 2.

The dependent measures were the same as those used in Experiment 1. These included braking response measures used in the Experiment 1 and the subjective ratings of urgency, annoyance, appropriateness, and the NASA TLX measures of workload. As in the first experiment, we also included several safety-related measures such as the number of collisions and the relative velocity at the point of impact.

3.1.4. *Experimental design and protocol*: The reduced complexity of the design made it possible to shorten the experiment. Experiment 2 lasted approximately 1.5 h. The participants drove four scenarios, one with each of the collision warning sound parameter combinations. Each scenario had four collision warnings (two hits, two false alarms) and four e-mail alerts (two concurrent, two non-concurrent with the collision warning). The warning algorithm used was the same as in Experiment 1. The scenery and ambient traffic were identical to Experiment 1. The four experimental scenarios were all 29.5 km long. The practice scenario remained the same.

Table 3. Summary of the conditions in Experiment 2

Independent variable	Conditions
Collision warning accuracy (within)	Accurate identification of a collision situation (Hit) False alarm (FA)
E-mail alert timing (within)	1 s before the collision warning (present) Not concurrent with collision warning (absent)
Collision warning density (within)	High burst density (0.80) Low burst density (0.20)
Collision warning duration (within)	Long duration (2.26 s) Short duration (0.44 s)

The e-mail task was also identical to that used in Experiment 1. Again, participants were given a bonus as the end of the experiment if they responded to all e-mail messages appropriately. As in Experiment 1, participants completed subjective workload ratings and a questionnaire at the end of each scenario. After the experiment, the experimenter debriefed the participants and solicited general comments.

3.2. *Results*

3.2.1. *Rated urgency*: The sound parameters for the collision avoidance warning failed to significantly influence rated urgency. Burst density and warning duration both failed to have a statistically significant effect on the mean urgency ratings of the collision avoidance warnings, $F(1,15) = 2.04$, $p = 0.1741$ and $F(1,15) = 0.09$, $p = 0.7658$.

3.2.2. *Rated annoyance*: The effect of burst density on annoyance approached statistical significance, $F(1,15) = 4.24$, $p = 0.0573$. As expected, the high-density collision warning had a mean rating of 5.10 and was perceived to be more annoying than the low-density collision warning, which had a mean rating of 4.27. The warning duration did not have a statistically significant effect on rated annoyance, $F(1,15) = 2.54$, $p = 0.1317$, but the trend in the mean ratings was in the expected direction, with the long duration being perceived as more annoying (5.08) than the short duration (4.23).

3.2.3. *Rated appropriateness*: Warning density affected the perceived appropriateness of the collision warning, $F(1,15) = 10.81$, $p = 0.0050$. With a mean of 4.13, the high-density collision avoidance warning was perceived as less appropriate than the low-density warning, which had a mean of 5.61. Warning duration, however, did not affect perceived appropriateness, $F(1,15) = 0.70$, $p = 0.4153$. Figure 4 shows the subjective ratings for urgency, annoyance, and appropriateness for the collision warning according to changes in density. As in Experiment 1, annoyance and appropriateness ratings of the collision warning alerts were related, with a significant

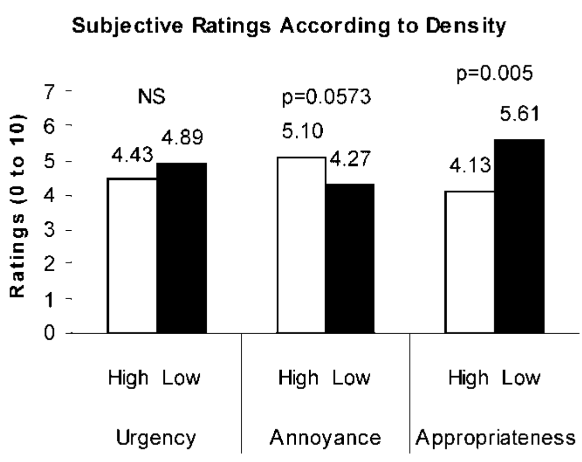

Figure 4. Subjective ratings of the collision warning according to density.

negative correlation of -0.253, $p < 0.01$. The significant correlation between urgency and appropriateness of the collision warning alerts was not replicated in this experiment and, in fact, the data show a slight negative correlation of -0.180, $p < 0.05$, with higher ratings of urgency associated with lower ratings of appropriateness.

3.2.4. *Subjective workload*: The analysis of the combined NASA TLX subjective ratings failed to show any effect of density or duration on subjective workload, $F(1,15) = 1.84$, $p = 0.1948$. As in Experiment 1, a significant correlation between the rated annoyance of the collision warning alert and subjective workload was found, 0.462, $p < 0.01$. The correlation between the rated urgency of the collision warning and the subjective workload was 0.009, $p > 0.05$. Collision warnings rated as more annoying were associated with higher levels of workload, but ratings of warning urgency were not associated with workload ratings. As in the first experiment, the urgency mapping principle suggests that larger differences between rated urgency between the collision warning and the e-mail alert will be associated with lower levels of subjective workload. The correlation between the difference in rated urgency of the collision warning and the e-mail alert and subjective workload was -0.218, $p > 0.05$. Unlike the first experiment, the statistical significance of this relationship suggests that appropriate urgency mapping can reduced workload. The correlation between workload and combined rated annoyance of the collision avoidance warning and e-mail alert was substantial, 0.524, $p < 0.01$. Similar to the results of the first experiment, more annoying warnings and alerts are associated with higher levels of subjective workload. Consistent with the first experiment, these results suggest that to reduce workload, urgency mapping may be less important than reducing the annoyance associated with warnings and alerts.

3.2.5. *Driving performance during lead vehicle braking events*: The data from all drivers were combined to form a database containing 119 imminent collision situations. Of the total 128 lead vehicle braking events, nine were excluded. Some of these were excluded because drivers were driving too slowly to experience the effect of the lead vehicle deceleration or lost the lead vehicle during a braking event. An additional 13 data points were not analyzed for the accelerator release reaction time because the drivers drove too fast. Events in which the lead vehicle braked are referred to as braking events.

The e-mail alert affected the accelerator release reaction time, $F(1,15) = 5.18$, $p = 0.0380$. Drivers responded to braking events in the presence of the e-mail alert responded 270 ms faster than they did when the e-mail alert was not present. Figure 5 shows the means for all conditions. Interestingly, neither the density nor the duration of the warning had an effect on accelerator release reaction time. An order effect was also seen in the accelerator release reaction time, $F(3, 36) = 3.00$, $p = 0.0430$. The braking events encountered in the first scenario resulted in a mean accelerator release time of 1.1 s. The second, third, and fourth scenarios resulted in mean accelerator release times of 0.62, 0.70, and 0.70 s, respectively. Unlike the results from the first experiment, there was no significant correlation between reaction time and the rated urgency of the collision warning.

Density significantly affected the accelerator to brake transition time, $F(1,15) = 5.35$, $p = 0.0353$. When drivers had the higher density collision avoidance warning they took 1.05 s to move their foot from the accelerator to the brake

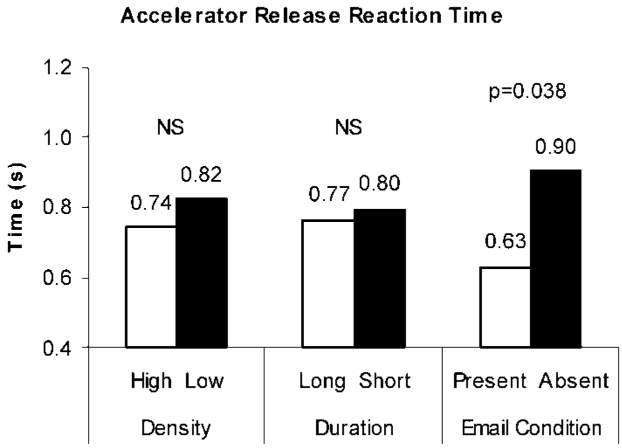

Figure 5. Mean accelerator release reaction times for density, duration, and e-mail condition during lead vehicle braking events.

compared to 0.79 s for the low-density warning. This result is similar to that seen for the accelerator to brake transition time in Experiment 1, where the high-urgency collision warning resulted in a faster accelerator release and a more controlled braking response characterized, in part, by a longer transition time; however, in this experiment density failed to have a statistically significant effect on accelerator release reaction time. One possible explanation is that the statistical power of the experiment was insufficient to detect an effect on accelerator release reaction time. This is consistent with the small F value for brake transition time and the shorter accelerator release reaction time in the high-density condition. Even if this explanation is correct, the small magnitude of this effect suggests that it may have little practical importance. There were no other significant effects for the accelerator to brake transition time.

An order effect was seen in the maximum deceleration level, $F(3,36) = 15.18$, $p < 0.0001$. The hit events experienced in the first scenario had a higher maximum deceleration (0.58 g) than those in the second, third, and fourth scenarios (0.47, 0.46, 0.46 g). The presence or absence of the e-mail system also affected the maximum deceleration level, $F(1,15) = 4.73$, $p = 0.046$. When the e-mail alert was present, drivers did not brake as hard as when the e-mail alert was absent. In other words, braking events in which the e-mail system was present resulted in a smaller maximum deceleration compared to events where the system was absent (0.47 vs. 0.51 g). An interaction between the collision avoidance warning urgency and e-mail condition was also seen for the maximum deceleration, $F(1,15) = 4.68$, $p = 0.0471$. When they were experiencing the long collision avoidance warning, the drivers braked to the same degree when the e-mail system was present or absent (0.49 g). However, when the drivers experienced the short collision avoidance warning, they braked harder when the e-mail system was absent versus when it was present (0.52 and 0.46 g, respectively). There were no significant differences found in the average deceleration levels achieved during the braking events.

The collision and collision velocity showed significant order results; all three collisions occurred in the first scenario ($F(3,36) = 5.50$, $p = 0.0032$ for collisions and $F(3,36) = 3.05$, $p = 0.0411$ for collision velocity).

3.3. *Discussion*

This experiment focused on two issues. First, it examined the effect of more extreme sound parameters on subjective ratings of urgency of the collision warning. This manipulation explored how well drivers are able to make absolute judgments of warning urgency. Second, it examined the effect of a longer interval between the e-mail alert and the collision warning. This manipulation revealed how the initial engagement in the e-mail interaction affected response to a collision warning.

3.3.1. *Effects of sound parameters on rated urgency, annoyance, and appropriateness*: As in the first experiment, sound parameters did not affect rated urgency. Neither the density nor the duration of the collision warnings influenced the urgency ratings. Because the warnings were drastically different with respect to density and duration, it seems that the failure to influence subjective ratings was not due to a failure to make the warnings distinctly different. In fact, the density values used here were the same ones used in previous studies that found very large differences in rated urgency when participants' only task was to rate the warnings. One contribution to this outcome, which is consistent with the results of the first experiment, is that people tend not to make precise absolute judgments of urgency, particularly in the context of multi-task performance. The inability to make direct comparisons, the demands of multi-task performance, and the delay between hearing and rating the warnings, seem to compromise drivers' ability to make reliable judgments of the urgency of different warning sounds. In addition, the strong effect of context may overwhelm the influence of sound parameters. Collision warnings tend to be rated as highly urgent and e-mail alerts tend to be rated as less urgent, independent of the particular sounds used in the warnings.

The perceived annoyance and appropriateness ratings also corresponded to the results of the first experiment, with the high-urgency collision warning being rated both more annoying and less appropriate than the low-urgency warning. In addition, the negative correlation between annoyance and appropriateness that was found in the first experiment was replicated. These results emphasize the importance of both subjective annoyance and appropriateness in warning evaluation. As in the first experiment, the sound parameters of the warnings affected driving performance, even though the subjective ratings indicate that drivers failed to differentiate the urgency of the warnings. The accelerator to brake movement times show some weak evidence that high-urgency warnings tended to produce more gradual response to a braking lead vehicle. Consistent with the urgency mapping principle, this experiment showed a modest negative association between the differences in rated urgency of the collision warning and e-mail alert and subjective workload. However, subjective workload was more strongly related to the combined rated annoyance of the collision warning and e-mail alert. These results are consistent with the first experiment and suggest that alert parameters affecting annoyance may have a more important effect on workload than those affecting urgency mapping.

3.3.2. *Effects of e-mail alerts and collision warnings on driving performance*: Contrary to the results of the first experiment, this study found that the e-mail alert facilitated braking performance. Increasing the delay time between the e-mail alert and the collision warning from 300 to 1000 ms reversed the alert's effect on driving performance, from delaying the braking response to speeding the braking response. With a 1000 ms delay, the e-mail alert induced a faster accelerator release. In addition, the other aspects of the response process, such as the accelerator to brake transition time, brake to maximum brake, and the mean and maximum deceleration, indicated a more controlled and less abrupt braking response when the alert preceded the warning. Contrary to the temporal interference predicted by the PRP effect, the e-mail alert speeded drivers' response to the collision alert. This may reflect strategic compensatory behaviour, where drivers prepare for the potential distraction of the e-mail interaction by removing their foot from the accelerator or preparing to remove their foot from the accelerator. Alternately, it may be an artifact of the simulator environment, in which drivers learned that one of three events would occur when the e-mail alert sounded. The e-mail alert might be followed by a collision warning and a braking event, by a collision warning and no braking event, or by nothing. If drivers learn to associate the e-mail alert with a braking lead vehicle then the e-mail alert may act as an early collision warning. Further investigation into potential learning effects found that as participants progressed through the experiment and experienced more braking events, the accelerator release reaction time decreased significantly, with an average release time of 1.10 s for events occurring in the first scenario, 0.63 s for the second scenario, 0.70 s for the third scenario, and 0.70 s for the fourth scenario, $F(3,36) = 3.00$, $p = 0.043$. However, this effect did not depend on the presence or absence of the e-mail alert, $F(3,33) = 0.98$, $p = 0.4208$, arguing against the idea that drivers learned to associate the e-mail alert with a collision situation. Therefore, it seems more likely that drivers compensated for the potential distraction of the e-mail task by preparing for a potential braking event. Drivers do not always act as passive recipients of driving and in-vehicle task demands, but sometimes adjust their behaviour in anticipation of high-demand periods.

4. Conclusions

There are several caveats in generalizing the results of this simulator study. The study exposed drivers to warnings for only a short period of time, and so the results may not predict long-term response to the warnings. It is possible that long-term exposure to the warnings would yield different results. In an effort to simulate the worst case situation of future collision warning systems, drivers experienced a somewhat unrealistic level of 50% false alarms that occurred randomly without any relationship to the driving environment. In a recent review of trust and its effects on reliance Lee and See (2004) suggest that false warnings would be less problematic to the extent that the false warnings are predictable or if drivers can adopt a strategy that can benefit from the imperfect information they provide (Muir and Moray 1996). It would be useful to explore how the type of failure influences driver attitudes and response to warnings. In addition, it may be improbable that 16 braking events would occur in a 3 h time period of driving. This high frequency of collision warnings might have promoted greater attentiveness than is typical in routine driving. Long-term adaptation may be different than that observed in the simulator; however, these results provide initial indicators of what factors might merit more extensive exploration in a longer-term on-road study.

Multiple resource theory (Wickens 2002) and the bottleneck theory (Pashler 1994) are two popular descriptions of human performance decrements in multi-task situations. Multiple resource theory predicts decrements associated with competition for shared resources. In these experiments, this competition occurred when the collision warning and e-mail alert both demand auditory perceptual resources.

Response occurred with different modalities (vocal and manual), and so an e-mail task that required a manual response might generate performance decrements. The bottleneck hypothesis predicts performance decrements associated with response selection associated with collision avoidance and e-mail acknowledgement. The bottleneck theory would predict greater interference for the 300 ms delay between the e-mail alert and the collision warning compared to the 1000 ms delay. The consequence of the multi-task demands should be reflected in subjective workload and reaction time to the lead vehicle braking events. The data were not completely consistent with these theories. In the first experiment the reaction time to the lead vehicle increases with the high-urgency e-mail alert, but not with the low-urgency alert. In second experiment, response to the lead vehicle actually improves when the e-mail alerts occur during the braking events. At least some drivers seemed to respond to the e-mail alert by anticipating the demands of the e-mail task and protecting the driving task. These results suggest that theories of multi-task performance must consider how warnings influence driver attitudes and how warnings influence the strategic responses of drivers as they adapt their driving in anticipation of high-demand situations (Adams *et al*. 1991, Moray *et al*. 1991, Raby and Wickens 1994).

4.1. *Implications for design*

These experiments show that sound parameters associated with increased urgency can enhance braking response to collision warnings. Designing collision avoidance warnings that are very urgent may reduce driver reaction time to warnings in a simulator setting, but the trade-off is that the resulting warning may annoy the driver and undermine system acceptance and eventually comprise driver response to the system in actual driving situations. Drivers who feel that an alert is inappropriate for the situation or annoying may reject the system. The results suggest that warning design should consider how sound parameters affect urgency, annoyance, and appropriateness. Specifically, the second experiment showed that the longer, high-urgency collision warnings did not provide a statistically significant reduction in drivers' reaction time to collision situations. Thus, it may be possible to use the less annoying, shorter warnings without compromising driving performance, while enhancing acceptance. Considering the relationship between annoyance and appropriateness may be particularly useful in designing alerts that are both effective and accepted.

Interestingly, subjective workload in these experiments did not respond according to Edworthy's urgency mapping principle. According to this principle, a high-urgency collision warning and a low-urgency e-mail alert should combine to reduce the ambiguity of the warning and alert and lead to lower workload. Conversely, a low-urgency collision warning and high-urgency e-mail alert combination led to higher levels of workload. The difference in rated urgency for the e-mail alert and the collision warning should have been negatively correlated with subjective workload—only the second experiment showed a weak negative correlation. In contrast, the sum of the rated annoyance of the e-mail alert and the collision warning had a relatively strong positive correlation with subjective workload in both experiments. These relationships suggest that there is an important trade-off associated with increasing the urgency of warnings to enhance response time and minimizing the annoyance associated with highly urgent warnings. For this reason, designers should consider trading off the benefits of a highly urgent warning with the costs associated with greater annoyance. The relative importance of the annoyance trade-off depends on the types of alerts and the frequency of false warnings. Urgency mapping and the effect of the warning on performance is clearly most important for highly critical warnings that have few false alarms. Alerts that are less critical or have more false alarms might benefit from a greater emphasis on the annoyance trade-off. In either case, annoyance trade-off analysis should complement urgency mapping in warning design.

Not surprisingly, overlapping two warnings can impair driver response to lead vehicle braking events, but only for certain interstimulus intervals. Drivers' reaction to a braking lead vehicle was degraded when an e-mail alert preceded the collision warning by 300 ms, but was enhanced when the alert preceded the collision warning by 1000 ms. The dynamics of the in-vehicle task and the associated distraction depend on more than the timing of the initial alert onset and so defining a specific interstimulus interval and extrapolation of these results requires caution. However, the results are encouraging. These findings suggest eliminating e-mail alerts from somewhere between 300 and 1000 ms before the onset of the collision warning could help reduce dual task conflict that might arise from the joint demands of collision warnings and in-vehicle information systems. The same algorithm used for the collision warning system, but with a more conservative parameter setting could lock out e-mail alerts whenever there is evidence for an emerging collision situation. The low cost of a delayed e-mail alert makes this solution quite feasible even if the system is imperfect. Using the same strategy to block telephone calls or to interrupt ongoing telephone calls might not be acceptable because early detection of emerging collision situations are more susceptible to false alarms. Drivers are unlikely to accept a system that periodically ends their telephone calls for no apparent reason, but might be quite happy to have a system that delays e-mail notifications until the driver reaches a low-demand situation. More generally, the lack of any multitask performance decrement with long interstimulus intervals suggests drivers can strategically adjust their driving in anticipation of high-demand situations when the interstimulus interval gives them the opportunity. Designers should consider how in-vehicle infotainment devices and collision warning systems affect drivers' ability to anticipate and respond to high demand situations.

4.2. *Implications for evaluation*

Studies investigating the effects of sound parameters on participants' perceptions typically involve a rapid succession of different warnings, which makes it possible to immediately rate each warning in comparison to others. In this study, the participants heard only one type of collision avoidance warning and one type of e-mail alert per drive and rated them at the end of the 12 min drive. In this situation, drivers were not able to provide precise and reliable ratings of urgency. Ratings of urgency and annoyance that are made immediately after hearing a warning may differ substantially from those made after a delay, particularly when this delay involves multi-task demands. This result has important implications for assessing auditory alerts and warnings. Evaluation in laboratory setting, simulator, and long-term exposure in on-road evaluations may all produce different ratings of urgency and annoyance. The difficulty in judging absolute levels of urgency suggests that pairwise comparison of alerts in a laboratory setting, where participants make relative judgments of many sounds, might be more precise than ratings in a simulator setting. These results also highlight the importance of context in interpreting sounds.

Collision warnings may tend to be rated as highly urgent independent of the sound parameters compared to e-mail alerts.

Acknowledgements

This paper presents research conducted in collaboration with Battelle's Human Factors Transportation Centre under contract number DTFH61-97-C-00061 to the Federal Highway Administration. Nazemeh Sobhi was the government technical representative for the contract until January 1999 and Tom Granda is the current government technical representative. The paper was greatly improved by the comments of two anonymous reviewers and by those of R. Stammers.

References

Adams, M. J., Tenney, Y. J. and Pew, R. W. 1991, State of the Art Report. Strategic workload and the cognitive management of advanced multi-task systems, CSERIAC 91-6, Crew Systems Ergonomics Information Analysis Center (CSERIAC) (OH: Wright-Patterson AFB).

Burgett, A. L., Carter, A., Miller, R. J., Najm, W. G. and Smith, D. L. 1998, A collision warning algorithm for rear-end collisions, 98-S2-P-31 (Washington, DC: National Highway Traffic Safety Administration).

Edworthy, J. 1994, The design and implementation of nonverbal auditory warnings, *Applied Ergonomics*, **25(4)**, 202–210.

Edworthy, J. and adams, A. 1996, *Warning Design: A Research Perspective* (Bristol, PA: Taylor & Francis).

Edworthy, J., Hellier, E. and Hards, R. 1995, The semantic associations of acoustic parameters commonly used in the design of auditory information and warning signals, *Ergonomics*, **38(11)**, 2341–2361.

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

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

Hart, S. G. and Staveland, L. E. 1988, Development of NASA-TLX (Task Load Index): Results of experimental and theoretical research, in P. A. Hancock and N. Meshkati (eds), *Human Mental Workload* (Amsterdam: North Holland), 139–183.

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

Kiefer, R., Leblanc, D., Palmer, M., Salinger, J., Deering, R. and Shulman, M. 1999, Development and validation of functional definitions and evaluation procedures for collision warning/avoidance systems, DOT HS 808 964 (Washington DC: Crash Avoidance Metrics Partnership).

Lee, J. D., McGehee, D. V., Brown, T. L. and Reyes, M. L. 2002, Collision warning timing, driver distraction, and driver response to imminent rear end collisions in a high-fidelity driving simulator, *Human Factors*, **44(2)**, 314–334.

Lee, J. D. and See, K. A. 2004, Trust in technology: Designing for appropriate reliance, *Human Factors*, in press.

Lerner, N. 1991, Multiple attribute evaluation of auditory warning signals for in-vehicle crash warning systems, DTNH22-91-C-07004 (Washington, DC: National Highway Transportation Safety Administration).

Marshall, D., Lee, J. D. and Austria, A. 2001, Annoyance and urgency of auditory alerts for in-vehicle information systems, *Human Factors and Ergonomics Society Annual Meeting*, **2,** 1627–1631.

Meredith, C. and Edworthy, J. 1995, Are there too many alarms in the intensive-care unit: An overview of the problems, *Journal of Advanced Nursing*, **21(1)**, 15–20.

Moray, N., Dessouky, M., Kijowski, B. A. and Adapathya, R. 1991, Strategic behavior, workload, and performance in task scheduling, *Human Factors*, **33(6)**, 607 – 629.

Muir, B. M. and Moray, N. 1996, Trust in automation 2: Experimental studies of trust and human intervention in a process control simulation, *Ergonomics*, **39(3)**, 429 – 460.

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

Raby, M. and Wickens, C. D. 1994, Strategic workload management and decision biases in aviation, *The International Journal of Aviation Psychology*, **4(3)**, 211 – 240.

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

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