Transportation Research Part F 47 (2017) 23–33

Contents lists available at ScienceDirect

Transportation Research Part F

journal homepage: www.elsevier.com/locate/trf

Considering visual-manual tasks performed during highway driving in the context of two different sets of guidelines for embedded in-vehicle electronic systems

David G. Kidd[a,*], Jonathan Dobres[b], Ian Reagan[a], Bruce Mehler[b], Bryan Reimer[b]

[a] *Insurance Institute for Highway Safety, 1005 N. Glebe Road, Arlington, VA 22201, USA*
[b] *MIT AgeLab & New England University Transportation Center, 77 Massachusetts Avenue, E40-291, Cambridge, MA 02139, USA*

A R T I C L E I N F O

Article history:
Received 2 December 2015
Received in revised form 8 March 2016
Accepted 3 April 2017

Keywords:
NHTSA driver distraction guidelines
Alliance driver focus guidelines
Visual demand assessment
Embedded vehicle system

A B S T R A C T

The Alliance of Automobile Manufacturers and the National Highway Traffic Safety Administration have each developed a set of guidelines intended to help developers of embedded in-vehicle systems minimize the visual demand placed on a driver interacting with the visual-manual interface of the system. Though based on similar precepts, the guidelines differ in the evaluation methodologies and the criteria used to define safe levels of visual demand. The current study compared the pass/fail conclusions from applying the two guidelines. Four visual-manual tasks were evaluated using two embedded in-vehicle systems (Volvo Sensus, Chevrolet MyLink) during highway driving. Only a preset radio tuning task met the threshold for acceptable visual demand in both guidelines. The pass/fail conclusions for three of the four tasks [manual radio tuning (fail), preset radio tuning (pass), easy contact calling (fail)] performed using either system were the same for both guidelines; calling a contact with multiple possible numbers using MyLink failed both guidelines, and with Sensus the task passed the Alliance guidelines but not NHTSA's. Exploratory analyses suggested that broadening the age range of the participant sample specified in the Alliance guidelines beyond 45–65 year olds did not change pass/fail conclusions. Results from a Monte Carlo simulation suggested that relying on data from a single trial per the NHTSA guidelines may reduce the repeatability of pass/fail conclusions. Interestingly, the manual radio tuning task failed to pass both sets of guidelines, even though the organizations used it as a reference task for setting acceptable levels of visual demand. Perhaps this indicates that radios have become more difficult to tune than the ones that provided the basis for the guidelines; however, naturalistic driving studies have not indicated increased risk from tuning more modern radios. Analysis of glance behavior during naturalistic driving may provide opportunities to further refine the acceptable thresholds for visual demand.

1. Introduction

Two organizations offer guidance to assist developers of embedded in-vehicle systems in minimizing the visual demand placed upon a driver interacting with the systems. Both organizations specify a set of principles and best practices and a

* Corresponding author.
E-mail address: dkidd@iihs.org (D.G. Kidd).

http://dx.doi.org/10.1016/j.trf.2017.04.002
1369-8478/

methodology for evaluating various tasks performed using the visual-manual interface of embedded in-vehicle systems to determine whether they should be permissible when the vehicle is moving, that is, represent an acceptable degree of visual demand and associated risk. However, the evaluation methodologies and the criteria used to define safe levels of visual demand differ. There is concern about how these differences affect the pass/fail conclusions from task evaluations intended to determine whether a task meets a threshold of visual demand deemed acceptable and debate as to the sensitivity of some measurement criteria.

Members of the Alliance of Automobile Manufacturers and other experts developed a set of principles, criteria, and evaluation procedures centered on driver interactions with the visual-manual interfaces of information and communication systems produced by original equipment manufacturers and aftermarket vendors (Alliance of Automobile Manufacturers, 2006; hereafter "the Alliance guidelines"). The Alliance guidelines include 24 principles to promote the safe design of human-machine interfaces that center on (1) installation, (2) the presentation of information, (3) system behavior, (4) driver interaction, and (5) instructions about system use. Principle 2.1 states that visual displays should be designed so that a desired task can be completed with sequential glances that are brief enough not to adversely affect driving.

The Alliance guidelines provide options for verifying compliance with Principle 2.1. Tasks are evaluated according to Principle 2.1 Alternative A using two criteria. Criterion A1 limits the duration of single glances to task-related controls and displays, and criterion A2 limits the total duration of all task-related glances to these areas. Both criteria use driver glance behavior when manually tuning a "traditional" radio to a new station on a different frequency band as a reference for an acceptable level of visual demand (i.e., the manual radio tuning task). The manual radio tuning task was selected to define an upper limit for visual interaction because of its long-standing use in research and "its impacts on driver eye glance behavior, vehicle control, and object-and-event detection are reasonably well understood" (p. 40).

One method for evaluating whether a task meets these criteria is to recruit an equal number of male and female drivers 45–65 years old who are unfamiliar with the device being tested and then measure the number and lengths of glances to the display(s) and task-related controls when the task is performed at least twice during daytime driving along a dry, divided highway with low to moderate traffic density and a posted speed limit of 45 mph or lower. The sample size must be large enough to produce statistically significant results. A task meets criterion A1 if the 85th percentile mean duration of glances to task-related controls or displays for the sample is less than or equal to 2.0 s. A task meets criterion A2 if the 85th percentile total duration of glances to task-related controls and displays for the sample is less than or equal to 20 s.

In 2013, the National Highway Traffic Safety Administration (NHTSA) published guidelines intended to reduce visual-manual driver distraction during interactions with the visual-manual interfaces of integrated original equipment communication and information systems (National Highway Traffic Safety Administration, 2013; hereafter "the NHTSA guidelines"). The guidelines are applicable to non-driving-related tasks including radio tuning as well as driving-related tasks that are not related to the safe operation and control of the vehicle (e.g., navigation). The NHTSA guidelines adopted many of the principles in the Alliance guidelines; for example, the mounting location must be easy to see and reach, the maximum downward viewing angle is 30 degrees, and the driver should control the pace of task interactions not the system. However, the NHTSA guidelines deviate with regard to (1) the criteria used to determine if a task is acceptable to be performed during driving, (2) the sampling requirements, (3) the testing environment, and (4) other aspects of the evaluation methodology. Some of the differences, such as NHTSA specifying measurement of glances off the forward roadway as opposed to quantifying task-specific related glances, were characterized as attempts to simplify testing procedures. Categorizing glances to "on the forward roadway" and "off the forward roadway" was seen as reducing some of the challenges in using automated eye tracking equipment for glance measurement.

Measuring off-road eye glance behavior during simulated driving is one method used for "acceptance testing" of a task per the National Highway Traffic Safety Administration (2013) guidelines. This approach requires a sample of 12 men and 12 women evenly distributed across four age groups (18–24, 25–39, 40–54, and 55+ years older). The participants are instructed to keep their vehicle 220 feet behind a lead car while performing tasks on a straight, four-lane undivided roadway with a posted speed limit of 50 mph. Glances away from the forward road scene during a single trial are measured to determine if a task meets the following three criteria: no more than 15 percent of 21 of the 24 participants' glances away from the forward road scene are longer than 2 s, the average duration of a glance away from the forward road scene is less than or equal to 2 s for 21 participants, and the total duration of glances away from the forward road scene is less than or equal to 12 s for 21 participants.

Consistent with the Alliance guidelines, the three criteria that tasks must meet in the NHTSA guidelines are based on the level of visual demand associated with manual radio tuning. However, the NHTSA guidelines have been criticized for specifying a more conservative threshold for the total duration of glances away from the roadway (12 s) than the Alliance permits for glances to the system or related controls (20 s). Recent research on glance behavior measured during manual radio tuning while driving on a simulated roadway and a test track was used to support the lower threshold (Perez et al., 2013). However, NHTSA's interpretation of these data has been questioned because a test track setting is not used for task evaluation in the NHTSA guidelines (Young, 2015) and the data may not adequately represent manual tuning performance across all radios (Alliance of Automobile Manufacturers, 2012).

Some other notable differences between the Alliance and NHTSA guidelines could also affect the pass/fail conclusion for a given task. First, the Alliance guidelines focus on task performance by drivers ages 45–65 because it is believed that these drivers will perform worse, on average, than their younger counterparts (General Motors, 2012) whereas the NHTSA guidelines evaluate task performance across the full adult age range. Second, the Alliance guidelines specify that performance is evaluated across multiple task trials and do not exclude trials with errors, whereas performance from a single error-free trial is evaluated per the NHTSA guidelines. The use of multiple trials can improve the precision of measurements. Error-free trials are qualitatively different from trials with errors and could reasonably be expected to affect, at a minimum, task completion time (e.g., Reimer, Mehler, Dobres, & Coughlin, 2013). Third, as noted above, the Alliance guidelines focus on glances to the system or related controls, but the NHTSA guidelines are concerned with all glances away from the forward road scene, including those to driving-relevant areas like vehicle mirrors. If all other factors remained constant, this distinction should lead to consistently greater values for the total off-the-forward-roadway glance time than glances at interfaces. Moreover, the driving tasks in the guidelines also differ; the Alliance guidelines specify that testing include other vehicular traffic and specifies lower speeds than the simulated driving scenario defined in the NHTSA guidelines. Note that the Alliance guidelines also permit task evaluation using an instrumented vehicle on a test track or on actual roads.

Both sets of guidelines share the goal of identifying tasks that invoke a potentially unsafe level of visual demand on the driver. However, the different evaluation methodologies and pass/fail thresholds may lead to inconsistent determinations regarding the acceptability of performing a given task while driving. The purpose of this study was to compare the pass/fail conclusions from applying the Alliance and NHTSA task evaluation methodologies to four visual-manual tasks performed using two different production in-vehicle infotainment systems during highway driving. Exploratory analyses also were performed to evaluate basic assumptions made by the guidelines regarding the reliability of pass/fail conclusions and the composition of the sample population.

2. Method

The data come from an experimental study that examined the attentional demands of using a visual-manual or voice interface to perform various tasks during highway driving (Mehler et al., 2015). The sample consisted of 80 participants evenly divided among four age groups (20–24, 25–39, 40–54, 55–69); note that recruitment did not include 18 and 19 year-olds in the NHTSA-specified 18–24 age grouping. There was an equal number of men and women in each age group.

Half of the 80 participants were randomly assigned to drive a 2013 Volvo XC60, and the other 40 participants drove a 2013 Chevrolet Equinox. The XC60 was equipped with an embedded Sensus infotainment system (Fig. 1A), and the Equinox was equipped with an embedded MyLink infotainment system (Fig. 1B). Participants used each system to place calls to contacts stored in a connected Smartphone's phone book, enter destination addresses into the navigation system, and manually select specified radio stations.

Only data from tasks where participants interacted with the embedded system's visual-manual interface were included in the current analysis because these tasks are subject to both the Alliance and NHTSA guidelines. The tasks were calling a contact associated with a single telephone number (easy contact calling), calling a contact associated with multiple telephone numbers requiring additional input (hard contact calling), tuning the radio to a new station using a radio preset button (preset radio tuning), and manually tuning the radio to a new station on a different frequency band (manual radio tuning). The number and type of manual inputs required to complete each task as instructed during task training are shown in Table 1.

With the exception of preset radio tuning, participants always began each task on a general menu screen, so the first step was to access the radio or phone function. The easy and hard contact calling tasks required participants to search through a

A B

Fig. 1. (A) Chevrolet MyLink visual-manual interface and (B) Volvo Sensus visual-manual interface.

Table 1
Number (types) of manual inputs required to complete each visual-manual task as trained.

System	Task			
	Easy contact calling	Hard contact calling	Preset radio tuning	Manual radio tuning
Chevrolet MyLink	9 inputs (5 push-button, 4 rotary knob)	10 inputs (5 push-button, 5 rotary knob)	1 input (1 push-button)	3 inputs (2 push-button, 1 rotary knob)
Volvo Sensus	3 inputs (2 push-button, 1 rotary knob)	5 inputs (3 push-button, 2 rotary knob)	1 input (1 push-button)	5 inputs (3 push-button, 2 rotary knob)

Note: Rotating the rotary knob to scroll through a menu list was considered 1 input.

list of 108 phone contacts stored on a Smartphone connected to the infotainment system. Both systems required participants to first access the phone book of the Smartphone. Participants pressed the "TEL" button to access the phone book with the Sensus. Participants using MyLink used the rotary knob to highlight the phone function listed on the "Home" screen, and then pressed the "Menu/Sel" push-button on the rotary knob to select it. Then the rotary knob was used to highlight the "Phone book" option from a list, and select using the "Menu/Sel" push-button on the rotary knob.

The systems had different methods for searching the contact list. The Volvo Sensus required participants to use the rotary knob to search through the entire contact list for the target contact, and push the "OK" push-button on the rotary knob to select the highlighted contact. In contrast, the Chevrolet MyLink required participants to use the rotary knob and "Menu/Sel" push-button to highlight and select the alphanumeric bin (e.g., ABC, DEF) corresponding to the contact name, highlight and select the target contact from a shortened list of 18 contacts presented on the next menu screen, and then select the single number listed for the contact. In the hard contact calling task, participants were presented with multiple possible telephone numbers after selecting the target contact. Participants highlighted and selected the appropriate number (e.g., work, mobile) from a list using the rotary knob and integrated push-button with both systems.

The preset radio tuning task was completed by pressing a single push-button in the center console to select the appropriate preset station. The manual radio tuning task was consistent with the specifications for individual trials of a reference task described in the Alliance guidelines and adapted by NHTSA (2012, 2013). Following the NHTSA guidelines, the radio was playing at the start of the task but it was not the active function. Participants were instructed to tune the radio to a target frequency on a different frequency band approximately one-third of the band away from the starting frequency (e.g., AM 1030–1470 or FM 107.9–100.7).

Manual radio tuning with Sensus always began at the MyCar menu. Participants pressed the "Radio" push-button to activate the radio and then pressed it again to bring up a list of possible frequency bands. Participants turned the rotary knob to the target frequency band and pressed the "OK" push-button on the rotary knob to confirm the selection. Lastly, participants turned the rotary knob until the target radio frequency was reached.

Manual radio tuning using MyLink began at the "Home" screen. Participants pressed the "Source" push-button to bring up a list of possible frequency bands and then pressed it again to select the target band. Next, participants rotated the "Menu/Sel" rotary knob until reaching the target frequency.

All participants were trained on how to perform each task using the push-buttons and knobs on the center console with the vehicle parked. As specified by both guidelines, participants were encouraged to repeat the tasks until they felt comfortable. With the exception of the manual radio tuning task, participants were not required to complete the tasks in the manner in which they were trained if they discovered another method of completing the task during the experimental drive that was preferred. After training, participants began the experimental drive and performed the tasks while driving on highways (I-93 and I-495) in and around the greater Boston area. The experimental drive consisted of a divided interstate with three travel lanes in each direction largely surrounded by forest. Each lane was 15 feet wide, and the posted speed limit was 65 mph. The experimental drive began with a 30 min driving adaptation period beginning with 10 min of urban driving from Cambridge, Massachusetts to interstate highway I-93 followed by 20 min of driving on I-93 to interstate I-495. Participants performed the easy and hard contact calling tasks while driving on I-495, and performed the preset and manual radio tuning tasks on I-93 when returning to Cambridge.

Each participant performed each task twice. The two easy contact calling task trials always preceded the two hard contact calling task trials. Both trials of preset radio tuning and of manual radio tuning were completed after the contact calling tasks. Trial instruction and stimuli were delivered using pre-recorded audio prompts. Participants were instructed to give priority to driving safely even if meant delaying the start of a task or skipping the task entirely. A research associate seated on the right side of the second row of the vehicle provided supplemental instruction and guidance as needed. Participants who experienced unsafe weather or roadway congestion leading to unstable traffic flow where the posted speed could not be maintained were not included in the analyses.

2.1. NHTSA guidelines sample

A random sample of 24 participants from each vehicle sample of 40 participants was selected with the constraint that there were three men and three women from each age group (20–24, 25–39, 40–54, and 55–69). NHTSA's protocol for measuring eye glances during simulated driving only considers performance from a single, "error-free" trial following a practice trial. Accordingly, the second trial of each task was sampled if it was performed without error. If an error occurred, then the first trial was used if it was error-free and, if not, the participant was replaced.

2.2. Alliance guidelines sample

Data from participants between the ages of 45 and 65 formed the sample for the Alliance guidelines. The Alliance guidelines state that each task should be performed at least twice but do not specify an acceptable level of task performance (error rates, amount of backtracking, etc.). Participants who successfully completed the task with or without errors in both trials were included in the sample, and the dependent measures were averaged across both trials. A total of 12 participants (six women, six men) from each vehicle sample satisfied the age and performance criteria above and were selected. The Alliance guidelines recommend recruiting enough participants to sufficiently control for Type I and Type II error. An *a priori* power analysis performed using G*Power (Version 3.1.9.2; Faul, Erdfelder, Lang, & Buchner, 2007) indicated that a sample size of 12 provided adequate statistical power ($1 - \beta = 0.83$) to reject the null hypothesis that there is no difference between the mean value for the sample and a constant criterion value using a one-tailed *t*-test with a large effect size ($d = 0.8$) and alpha set to 0.05.

2.3. Dependent measures

Eye glance data were manually double-coded from video recordings of the driver's face per the International Organization for Standardization (ISO) standards 15007-1 (International Organization for Standardization, 2002) and 15007-2 (International Organization for Standardization, 2001). Discrepancies between the coders in glance location or timing were resolved by a third researcher (see Mehler et al. (2015) for details). Glance duration was defined as the first video frame where the eye began moving to a new location until the last video frame before the next eye movement began. In the analysis of the NHTSA guidelines, eye glances that were not to the forward roadway were categorized as being off-road, including driving-related glances (e.g., those to the vehicle mirrors). Three visual glance metrics were calculated for each task trial: (1) mean off-road glance duration, (2) percentage of the participant's off-road glances that were longer than 2 s, and (3) the total duration of off-road glances. The Alliance guidelines are based on glances to device-related areas. In these analyses, all glances to the center console were categorized as glances to the device and used to calculate the mean glance duration to the device and the total duration of glances to the device. Note that glances to task-related areas (e.g., buttons on the steering wheel) other than the center console were not considered as per the Alliance guidelines.

3. Results

3.1. NHTSA guidelines pass/fail conclusions

The NHTSA guidelines state that a task is suitable to be performed while driving if (1) 21 of the 24 participants' mean off-road glance duration is less than or equal to 2 s, (2) 85 percent of the off-road glances made by 21 participants during the data trial are not longer than 2 s, and (3) the sum of off-road glances during the trial is less than or equal to 12 s for 21 participants. Twenty-one of 24 participants corresponds to the 87.5th percentile value of the sample and was computed for each measure. The 87.5th percentile value for average off-road glance duration was less than 2 s for all four tasks in both vehicles (Table 2). The 87.5th percentile value for the percentage of a participant's off-road glances longer than 2 s was less than 15 percent for every task except for the hard contact calling task using the Volvo Sensus. Finally, the preset radio tuning task was the only task in both vehicles where the 87.5th percentile value for total duration of off-road glances was less than or equal to 12 s. Hence, according to the NHTSA guidelines, the only task determined to be acceptable for performing while driving using the Chevrolet MyLink or Volvo Sensus was preset radio tuning.

3.2. Alliance guidelines pass/fail conclusions

The Alliance guidelines state that a task is suitable to be performed while driving if (1) the 85th percentile average duration of a glance to the device for the sample is less than or equal to 2 s and (2) the 85th percentile of the average total time spent glancing to the device for the sample is less than or equal to 20 s. The 85th percentile value for the average glance duration to the device was less than or equal to 2 s for all four tasks performed using both the Chevrolet MyLink and Volvo Sensus (Table 3). The 85th percentile value for total glance duration to the device during preset radio tuning was less than 20 s for participants who used MyLink and participants who used Sensus. The 85th percentile value for total glance duration to the device also was less than 20 s for participants who used MyLink to perform the hard contact calling task, but not for participants who used Sensus. The total duration of glances to the device during manual radio tuning and easy contact calling exceeded 20 s for both MyLink and Sensus. Thus, the only tasks determined to be acceptable for performing while driving per the Alliance guidelines were preset radio tuning using the Chevrolet MyLink or Volvo Sensus and hard contact calling using MyLink.

Table 2
NHTSA guidelines task evaluation results for each task.

System	Task	Average off-road glance duration (87.5th percentile value)	Percent of participant's off-road glances longer than 2 s (87.5th percentile value)	Total duration of off-road glances (87.5th percentile value)	Pass/fail
Chevrolet MyLink	Manual radio tuning	1.22	9.09	**20.6**	Fail
	Preset radio tuning	1.00	0	2.76	Pass
	Easy contact calling	1.09	8.00	**21.36**	Fail
	Hard contact calling	1.08	0	**17.47**	Fail
Volvo Sensus	Manual radio tuning	1.28	11.76	**24.25**	Fail
	Preset radio tuning	1.24	0	3.71	Pass
	Easy contact calling	1.17	12.50	**12.09**	Fail
	Hard contact calling	1.36	**22.22**	**14.16**	Fail

Note: Bold cells indicate the task did not meet the criterion.

Table 3
Alliance guideline task evaluation results for each task.

System	Task	Average duration of glance to device (85th percentile value)	Total duration of glances to device (85th percentile value)	Pass/fail
Chevrolet My Link	Manual radio tuning	1.24	**23.7**	Fail
	Preset radio tuning	1.16	3.54	Pass
	Easy contact calling	1.09	**32.7**	Fail
	Hard contact calling	1.34	16.35	Pass
Volvo Sensus	Manual radio tuning	1.3	**30.7**	Fail
	Preset radio tuning	1.32	9.04	Pass
	Easy contact calling	1.2	**25.53**	Fail
	Hard contact calling	1.43	**26.46**	Fail

Note: Bold cells indicate the task did not meet the criterion.

3.3. Monte Carlo analysis of NHTSA pass/fail conclusions

Only 24 of the 40 participants assigned to each vehicle were used to evaluate each task according to the NHTSA guidelines, so there was an opportunity to draw multiple new samples with a different combination of participants to evaluate the consistency of the pass/fail conclusions for all four tasks. The same analysis could not be performed using the Alliance guidelines since all eligible participants were included in the Alliance guidelines sample. Six drivers in each of the four age groups were randomly sampled from the 40 participants assigned to each vehicle. Participants were drawn with replacement to allow for an even wider variety of participant combinations. A total of 1000 random samples of participants for each vehicle were drawn, and each sample's performance on the four tasks was evaluated per the NHTSA guidelines.

Table 4 shows the percentage of the 1000 samples in which the pass/fail conclusion matched the conclusion based on the original sample selected for each vehicle. All of the 1000 random samples had the same pass/fail conclusion as the original sample for manual radio tuning, preset radio tuning, and hard contact calling tasks using the Chevrolet MyLink, and all but four of the 1000 samples had the same conclusion for easy contact calling. Consistent with the original sample, manual radio tuning with the Volvo Sensus was determined to be unacceptable to be performed while driving for each of the 1000 random samples. However, the pass/fail conclusions for the other three tasks with the Volvo Sensus were less consistent. Among these three tasks, the pass/fail conclusions for 13.8, 25.6, and 3.0 percent of the 1000 random samples for the preset radio tuning, easy contact calling, and hard contact calling tasks, respectively, did not match the original sample.

Extreme values above a criterion's threshold could have generated inconsistent pass/fail conclusions in the Monte Carlo analysis. Boxplots were constructed to detect outliers greater than three times the inter-quartile range above the third quartile (Tukey, 1977). Across the three tasks without total agreement with the original sample, more outliers were observed for tasks performed by participants in the Volvo sample than for tasks performed by participants in the Chevrolet sample, and many more of these values exceeded the thresholds for the acceptable level of visual demand (Table 5). Outliers were most frequently observed for the percentage of off-road glances that were more than 2 s.

3.4. Age grouping assessment

The sampling requirements in the Alliance guidelines require participants to be between 45 and 65 years old; this is intended to provide a conservative evaluation of visual demand. Expanding the sample to include younger drivers should

Table 4
Percentage of 1000 random samples with the same pass/fail conclusion as the original sample. Original pass/fail conclusion indicated in parentheses.

Task	System	
	Chevrolet MyLink	Volvo Sensus
Manual radio tuning	100 (fail)	100 (fail)
Preset radio tuning	100 (pass)	86.2 (pass)
Easy contact calling	99.6 (fail)	74.4 (fail)
Hard contact calling	100 (fail)	97 (fail)

Table 5
Number of outlier values for each criterion and the number that exceed the criterion's threshold in parentheses.

System	Task	Mean off-road glance duration (outliers > 2 s)	Percent of participant's off-road glances longer than 2 s (outliers > 15%)	Total duration of off-road glances (outliers > 12 s)
Chevrolet MyLink	Manual radio tuning	0 (0)	0 (0)	0 (0)
	Preset radio tuning	0 (0)	8 (2)	1 (1)
	Easy contact calling	0 (0)	5 (0)	1 (1)
	Hard contact calling	0 (0)	5 (1)	0 (0)
Volvo Sensus	Manual radio tuning	0 (0)	3 (3)	2 (0)
	Preset radio tuning	0 (0)	2 (2)	0 (0)
	Easy contact calling	1 (1)	8 (2)	2 (2)
	Hard contact calling	1 (1)	9 (4)	1 (1)

result in similar or more favorable pass/fail conclusions. As an exploratory analysis, the Alliance guidelines were applied to the task performance of the 24 participants selected for the NHTSA sample.

Expanding the age range of the sample did not change the pass/fail conclusions for the four tasks. The 85th percentile values for the average duration of glances to the device during manual radio tuning (MyLink 1.18 s, Sensus 1.27 s), preset radio tuning (MyLink 1.16 s, Sensus 1.31 s), easy contact calling (MyLink 1.17 s, Sensus 1.17 s), and hard contact calling (MyLink 1.20 s, Sensus 1.41 s) were less than 2 s for both MyLink and Sensus. In both vehicles, the 85th percentile values for the total duration of glances to the device were greater than 20 s for the manual radio tuning (MyLink 22.71 s, Sensus 30.42 s) and easy contact calling tasks (MyLink 20.03 s, Sensus 23.53 s) and less than 20 s for the preset radio tuning task (MyLink 4.25 s, Sensus 9.04 s). The 85th percentile value for the total duration of glances to the device for the hard contact calling task was less than 20 s with MyLink (15.27 s) but more than 20 s with Sensus (21.48 s). The pass/fail conclusions for the four tasks when applying both guidelines and the Alliance guidelines to a driver sample with a broader age range are summarized in Table 6.

4. Discussion

The current study compared the pass/fail conclusions of the NHTSA and Alliance guidelines for four visual-manual secondary tasks performed using two embedded infotainment systems in production vehicles during highway driving. Both guidelines agreed that the visual demands of the manual radio tuning task and easy contact calling task were not acceptable to perform using either the Chevrolet MyLink or Volvo Sensus while driving under the testing conditions employed. Selecting a radio station using a push-button preset was acceptable with both MyLink and Sensus based on both guidelines. A hard contact calling task using Sensus was not acceptable based on either set of guidelines under the testing conditions employed; when using MyLink, the task passed the Alliance guidelines but not NHTSA's.

Table 6
Summary of the pass/fail conclusions.

System	Task	NHTSA guidelines	Alliance guidelines	Alliance guidelines with broader age range
Chevrolet MyLink	Manual radio tuning	Fail	Fail	Fail
	Preset radio tuning	Pass	Pass	Pass
	Easy contact calling	Fail	Fail	Fail
	Hard contact calling	Fail	Pass	Pass
Volvo Sensus	Manual radio tuning	Fail	Fail	Fail
	Preset radio tuning	Pass	Pass	Pass
	Easy contact calling	Fail	Fail	Fail
	Hard contact calling	Fail	Fail	Fail

Although the pass/fail conclusions across the four tasks were mostly consistent between the NHTSA and Alliance guidelines, there were some unexpected findings that suggest avenues for improving the robustness of the evaluation methodologies and different thresholds for an acceptable level of visual demand. Pass/fail conclusions with both guidelines are determined according to thresholds based on the visual demand of manual radio tuning. Interactions with the radio and other integrated vehicle controls were not associated with an increase in near-crash and crash risk among experienced drivers who participated in a naturalistic driving study (Klauer et al., 2014), supporting the rationale for using manual radio tuning as a benchmark for an acceptable level of visual demand. However, in the current study this task was found to have an unacceptable level of visual demand with both embedded systems during highway driving using both sets of guidelines.

Both the MyLink and Sensus systems had more functionality than the traditional radio used to determine the thresholds in the Alliance guidelines (Alliance of Automobile Manufacturers, 2006; pp. 46–48), but the basic features and the types of interactions required to tune the radio with both systems were similar. One difference was that the contemporary radios in the current study were tuned with a rotary knob rather than a dedicated tuning push-button as described in the Alliance guidelines; however, Perez et al. (2013) found that tuning a vehicle radio with push-buttons while driving on a closed test track led to greater visual demand and took longer to complete than tuning a radio with a rotary knob. Nevertheless, there are other differences between modern and traditional radios that may limit the implementation of manual radio tuning as a reference task in modern vehicles. For instance, the demand of touch screen responses may not be equivalent to the demand of push-buttons (e.g., Ranney, Baldwin, Parmer, Martin, & Mazzae, 2011). Similarly, the radio interfaces in many modern vehicles are not as distinct from other functions, especially in much more crowded center console areas, as those found in older vehicles with traditional radios.

The NHTSA and Alliance guidelines specify different driving tasks for task evaluation, and both differ from the driving task in the current study. The roadway in the current study consisted of an interstate highway with gentle curves, a 65 mph speed limit, and variable densities of free flowing traffic. The driving task in the NHTSA guidelines requires drivers to follow a lead car traveling at a slower speed of 50 mph along a simulated, straight roadway. The roadway described in the Alliance guidelines is somewhat similar to the roadway in the current study – a divided highway with low to moderate levels of traffic – but has a lower speed limit of 45 mph. The differences between these roadway environments may not be trivial considering that travel speed, roadway curvature, and traffic density affect the amount of driving task-relevant information in the visual scene and the visual demand of the driving task (Eby & Kostyniuk, 2004; Kujala, Mäkelä, Kotilainen, & Tokkonen, 2016). The differences between the testing environments (i.e., simulated or real-world) are potentially less important (Victor, Harbluk, & Engström, 2005; Wang et al., 2010).

Typical glances away from the roadway are about 1 s (Gellatly & Kleiss, 2000; Hoffman, Lee, McGehee, Macias, & Gellatly, 2005) and are rarely longer than 1.6 s (Victor et al., 2005; Wierwille, 1993), but drivers often change their behavior to cope with more visually demanding roadways. For example, they glance away from the roadway less frequently and for shorter periods of time on more visually demanding roadways (Senders, Kristofferson, Levison, Dietrich, & Ward, 1967; Tsimhoni & Green, 2001; Tsimhoni, Smith, & Green, 2004). More visually demanding roadways would therefore lead to fewer off-road glances lasting longer than 2.0 s that would exceed acceptance criteria specified in both guidelines. Drivers may also adjust their speed to control the amount of driving task-relevant information missed over time when looking away from the roadway (Kujala et al., 2016). Depending on the vehicle's speed, it is possible that glances shorter than 2.0 s could be unsafe in more demanding roadway environments like intersections. The dynamics and visual demands of real-world driving vary greatly, so the time-based acceptance criteria specified in the Alliance and NHTSA guidelines may only be relevant to a limited set of roadway environments that possess certain levels of visual demand. More generic acceptance criteria that incorporate the visual demands of the roadway and vehicle speed should be explored (e.g., Kujala & Mäkelä, 2015).

Both sets of guidelines allow tasks to be evaluated using the occlusion method. The occlusion method is a cost-effective way of measuring the visual demand of an in-vehicle task to inform interface design (e.g., Baumann, Keinath, Krems, & Bengler, 2004; Pettitt, Burnett, Bayer, & Stevens, 2004), but does not simulate driving task demand. In the occlusion method, the driver's visual sampling of the roadway and task interface are simulated by blocking the driver's view of the task interface for 1.5 s every 1.5 s (International Organization for Standardization, 2007). However, this does not impose any visual demands on the "driver" during a simulated glance to the roadway. Even having participants complete a simple visual-motor task when their view of the interface is blocked during an occlusion period can affect inferences made about the visual demand of a task (Monk & Kidd, 2007). Future research should compare the consistency of pass/fail conclusions using simulated driving tasks and the occlusion method and validate the predictive validity of both methods with performance during actual driving.

Some other notable differences in the task evaluation methods may affect the robustness of pass/fail conclusions within and between the approaches. For example, the driving simulator testing per the NHTSA guidelines is based on a single error-free trial, but the simulator, test track, or on-road testing per the Alliance guidelines considers task performance across multiple trials that may or may not include errors. Averaging performance across multiple trials should improve the precision of measurement and reduce some of the inconsistency of pass/fail conclusions observed for the Sensus system in the Monte Carlo analysis and noted in previous research (Aust, Rydström, Broström, & Victor. T., 2015). Interestingly, the NHTSA guidelines do not allow multiple data trials when tasks are evaluated in a driving simulator, but performance is averaged across five trials when using the occlusion method (NHTSA, 2013).

Both guidelines allow multiple tasks to be tested in a single session to improve testing efficiency; however, this approach may result in participant fatigue, confusion, or other problems (NHTSA, 2013). In the current study, the evaluation test results for both guidelines indicated that visual demand was greater for the easy contact calling task using MyLink than the hard contact calling task even though it had two fewer manual inputs. The easy contact calling task always preceded the hard contact calling task in the on-road study. Participants were thoroughly trained on every task before on-road testing began, but were unfamiliar with the system and may have needed to reacquaint themselves with the steps needed to call a contact. Any refamiliarization with the task steps would have affected performance during the easy contact calling task more than the hard contact calling task. Providing instructions, practice, and testing separately before each task, as directed by the NHTSA guidelines, could have minimized any order effect, but was not possible given the original study design (Mehler et al., 2015).

The current study used data from Mehler et al. (2015) who assessed the attentional demands of performing various tasks using the visual-manual and voice interfaces of MyLink or Sensus. In the experimental drive, Mehler et al. did not require participants to complete the contact calling tasks as instructed during training to allow for a more naturalistic assessment of the embedded interfaces. The Alliance guidelines do not address how to handle trials where the participant does not perform the task as instructed. However, task acceptance testing guideline E.10 in the NHTSA guidelines (pp. 41, NHTSA, 2013) indicates that an experimenter can invalidate and remove a trial where the participant was not judged to not genuinely attempt to perform the task as instructed. The NHTSA sample in the current study may have included trials where participants did not follow the sequence of steps they were instructed to perform for the contact calling task. Including these trials in the current study may limit the degree to which the NHTSA guidelines were applied appropriately; however, it may not have given that the experimenter is responsible for determining if such trials should be excluded.

The NHTSA guidelines also recruit a broader age range of drivers than the Alliance guidelines. Expanding the age range of drivers beyond that specified in the Alliance guidelines did not affect the pass/fail conclusions in the current study, but it did somewhat reduce the 85th percentile total duration of glances to the device for some tasks. For example, the 85th percentile total glance duration was almost 13 s less for the expanded age range of drivers compared with the original sample for the easy contact calling task using MyLink and about 5 s less for the hard contact calling task using Sensus. Relative to younger drivers, older drivers make glances longer than 2 s more frequently when interacting with an in-vehicle device and look away from the roadway for longer total periods of time (e.g., Wikman & Summala, 2005). These differences suggest that excluding younger drivers and limiting the driver sample to 45–64 year-olds as specified in the Alliance guidelines does provide a more conservative estimate of visual demand as intended.

Neither set of guidelines has verified the predictive validity of manual radio tuning as a benchmark for discriminating between tasks with levels of visual demand that do or do not compromise safety. On the other hand, naturalistic driving studies have linked certain glance behaviors to increased crash and near-crash risk. For example, glances lasting longer than 1.7 s have been associated with a significant increase in near-crash and crash risk (Liang, Lee, & Horrey, 2014). In addition to glance duration, the timing and distribution of off-road glances also are safety relevant. Drivers may miss sudden changes in traffic conditions even during short glances away from the roadway (Victor et al., 2015). Examinations of glance behavior in the 6 s surrounding crashes and near-crashes in two different naturalistic driving studies indicated that crash and near-crash risk was significantly greater if the total duration of off-road glances exceeded 2 s (Klauer, Guo, Sudweeks, & Dingus, 2010; Victor et al., 2015). Liang, Lee, and Yekhshatyan (2012) used naturalistic driving data to relate various methods for characterizing attention to the forward roadway with near-crash and crash risk. They found that the duration of the current off-road glance or summed duration of off-road glances during short time windows (e.g., 3 s) best explained the variance in near-crash and crash risk associated with driver glance behavior; measures that summarized glance behavior over longer periods of time were less precise. Based on these findings, restricting the total duration of off-road glances made during short time segments of task performance may be an improvement over restricting the total duration of off-road glances during the entire task.

5. Conclusions

The Alliance and NHTSA guidelines provide designers with a set of principles for limiting the visual demand of tasks performed using embedded in-vehicle systems and original equipment while driving, but they specify different criteria for determining if a task is safe to perform while driving. This study found that the two task evaluation approaches led to similar pass/fail conclusions when applied to four visual-manual tasks performed during highway driving. However, the visual demands of highway driving likely are different from the visual demands of the driving situations in both guidelines. Drivers often change their glance behavior or vehicle speed to cope with the visual demands of the roadway environment, and these behavioral changes may influence the consistency of pass/fail conclusions across roadway environments. Exploratory analyses of the data suggested that the NHTSA guidelines may produce inconsistent pass/fail conclusions with only one trial and would be improved by considering performance across multiple trials to minimize the influence of extreme outliers. The Alliance guidelines limit the age range of the testing sample, but broadening the sample age did not affect the pass/fail conclusions. Manual radio tuning was not acceptable when applying either set of guidelines, even though it is used as the benchmark for a safe level of visual demand. The Alliance and NHTSA pass/fail criteria were developed based upon detailed reviews of available empirical research around manual radio tuning. However, it can be argued that using a reference task to set acceptable levels of visual demand may no longer be the most scientific way to define thresholds. Understanding the association between visual distraction and risk of safety-relevant events has been facilitated by analysis of data from several naturalistic driving studies, and these data could help define appropriate thresholds for an acceptable level of visual demand.

Acknowledgments

The authors would like to thank Anne McCartt, Adrian Lund, David Zuby, and the other Institute staff involved with reviewing and commenting on this manuscript. This work was supported by the Insurance Institute for Highway Safety.

References

Alliance of Automobile Manufacturers (2006). Statement of principles, criteria and verification procedures on driver interactions with advanced in-vehicle information and communication systems: Including 2006 updated sections. Washington, DC. Available: <http://www.autoalliance.org/index.cfm?objectid=D6819130-B985-11E1-9E4C000C296BA163>. Accessed: Jun 26, 2015.

Alliance of Automobile Manufacturers (2012). Request for technical data and clarification to the National Highway Traffic Safety Administration concerning the Guidelines for reducing visual-manual driver distraction during interactions with integrated, in-vehicle, electronic devices. Docket Document No. NHTSA-2012-0053-0132. Washington, DC.

Aust, M. L., Rydström, A., Broström, R., & Victor. T. (2015). Better ways to calculate pass/fail criteria for the eye glance measurement using driving simulator test. In *Proceedings of the 4th international driver distraction and inattention conference*. Sydney, New South Wales: ARRB Group Ltd.

Baumann, M., Keinath, A., Krems, J. F., & Bengler, K. (2004). Evaluation of in-vehicle HMI using occlusion techniques: Experimental results and practical implications. *Applied Ergonomics, 35*(3), 197–205. http://dx.doi.org/10.1016/j.apergo.2003.11.011.

Eby, D. W. & Kostyniuk, L. P. (2004). SAfety VEhicles using adaptive Interface Technology (Task 2a). *Estimating driving task demand from crash probabilities: A review of the literature and assessment of crash databases*. Washington, DC: National Highway Traffic Safety Administration.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. http://dx.doi.org/10.3758/BF03193146.

Gellatly, A., & Kleiss, J. (2000). Visual attention demand evaluation of conventional and multifunction in-vehicle information systems. In *Proceedings of the human factors and ergonomics society 44th annual meeting* (pp. 3282–3285). Santa Monica, CA: Human Factors and Ergonomics Society.

General Motors (2012). Comment on the National Highway Traffic Safety Administration concerning the Guidelines for reducing visual-manual driver distraction during interactions with integrated, in-vehicle, electronic devices. Docket Document No. NHTSA-2010-0053-103. Washington, DC.

Hoffman, J., Lee, J., McGehee, D., Macias, M., & Gellatly, A. (2005). Visual sampling of In-Vehicle text messages: Effects of number of lines, page presentation, and message control. *Transportation Research Record, 1937*(1), 22–30. http://dx.doi.org/10.3141/1937-04.

International Organization for Standardization (2001). *Road vehicles: Measurement of driver visual behaviour with respect to transport information and control systems; Part 2: Equipment and procedures*. ISO 15007–2. Geneva, Switzerland.

International Organization for Standardization (2002). *Road vehicles: Measurement of driver visual behaviour with respect to transport information and control systems; Part 1: Definitions and parameters*. ISO 15007–1. Geneva, Switzerland.

International Organization for Standardization (2007). *Road vehicles: Ergonomic aspects of transport information and control systems; Occlusion method to assess visual demand due to the use of in-vehicle systems*. ISO 16673. Geneva, Switzerland.

Klauer, S. G., Guo, F., Sudweeks, J., & Dingus, T. A. (2010). *An analysis of driver inattention using a case-crossover approach on 100-car data*; Final report. Report No. DOT HS 811 334. Washington, DC: National Highway Traffic Safety Administration.

Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. *New England Journal of Medicine, 370*(1), 54–59. http://dx.doi.org/10.1056/NEJMsa1204142.

Kujala, T., & Mäkelä, J. (2015). Testing environment and verification procedure for in-car tasks with dynamic self-paced driving scenarios. In *Proceedings of the 4th International Driver Distraction and Inattention Conference*. Sydney, New South Wales: ARRB Group Ltd.

Kujala, T., Mäkelä, J., Kotilainen, I., & Tokkonen, T. (2016). The attentional demand of automobile driving revisited: Occlusion distance as a function of task-relevant event density in realistic driving scenarios. *Human Factors, 58*, 163–180.

Liang, Y., Lee, J. D.,& Horrey, W. J. (2014). A looming crisis: The distribution of off-road glance duration in moments leading up to crashes/near-crashes in naturalistic driving. In *Proceedings of the human factors and ergonomics society 58th annual meeting* (pp. 21012–2106). Santa Monica, CA: Human Factors and Ergonomics Society.

Liang, Y., Lee, J. D., & Yekhshatyan, L. (2012). How dangerous is looking away from the road? Algorithms predict crash risk from glance patterns in naturalistic driving. *Human Factors, 54*(6), 1104–1116. http://dx.doi.org/10.1177/0018720812446965.

Mehler, B., Kidd, D. G., Reimer, B., Reagan, I. J., Dobres, J., & McCartt, A. T. (2015). *Multi-modal assessment of on-road demand of voice and manual phone calling and voice navigation entry across two embedded vehicle systems*. Arlington, VA: Insurance Institute for Highway Safety.

Monk, C. A., & Kidd, D. G. (2007). R we fooling ourselves: Does the occlusion technique shortchange R estimates? In *Proceedings of the 4th international driving symposium on human factors in driver assessment, training, and vehicle design* (pp. 2–8). Iowa City, IA: University of Iowa.

National Highway Traffic Safety Administration (2012). *Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices*. Docket No. NHTSA-2010-0053. Washington, DC: U.S. Department of Transportation.

National Highway Traffic Safety Administration (2013). *Guidelines for reducing visual-manual driver distraction during interactions with integrated, in-vehicle, electronic devices: Version 1.01*. Docket No. NHTSA-2010-0053. Washington, DC: U.S. Department of Transportation.

Perez, M., Owens, J., Viita, D., Angell, A., Ranney, T. A., Baldwin, G.H.S., . . ., Mazzae, E. N. (2013). *Radio tuning effects on visual and driving performance – Simulator and test track studies*. Report No. DOT HS 811 781. Washington, DC: National Highway Traffic Safety Administration.

Pettitt, M. A., Burnett, G. E., Bayer, S., & Stevens, A. (2004). Assessment of the occlusion technique as a means for evaluating the distraction potential of driver support systems. *IEE Proceedings, Intelligent Transport Systems, 153*(4), 259–266. http://dx.doi.org/10.1049/ip-its:20060027.

Ranney, T. A., Baldwin, G. H. S., Parmer, E., Martin, J., & Mazzae, E. (2011). *Distraction effects of manual number and text entry while driving*. Report No. DOT HS 811 510. Washington, DC: National Highway Traffic Safety Administration.

Reimer, B., Mehler, B., Dobres, J., & Coughlin, J. F. (2013). *The effects of a production level "voice-command" interface on driver behavior: Reported workload, physiology, visual attention, and driving performance*. MIT AgeLab Technical Report No. 2013–17A. Cambridge, MA: Massachusetts Institute of Technology.

Senders, J., Kristofferson, A., Levison, W., Dietrich, C. W., & Ward, J. L. (1967). The attentional demand of automobile driving. *Highway Research Record, 195*, 15–33.

Tsimhoni, O. & Green, P. (2001). Visual demand of driving and the execution of display-intensive in-vehicle tasks. In *Proceedings of the human factors and ergonomics society 45th annual meeting* (pp. 1586–1590). Santa Monica, CA: Human Factors and Ergonomics Society.

Tsimhoni, O., Smith, D., & Green, P. (2004). Address entry while driving: Speech recognition versus a Touch-Screen keyboard. *Human Factors, 46*(4), 600–610. http://dx.doi.org/10.1518/hfes.46.4.600.56813.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, PA: Addison-Wesley.

Victor, T. W., Bärgman, J., Boda, C., Dozza, M., Engström, J.A., Flannagan, C., . . ., Markkula, G. (2015). *Analysis of naturalistic driving study data: safer glances, driver inattention, and crash risk*. SHRP2 Report S2–S08A-RW-1. Washington, DC: Transportation Research Board.

Victor, T. W., Harbluk, J. L., & Engström, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F, 8*(2), 167–190. http://dx.doi.org/10.1016/j.trf.2005.04.014.
Wang, Y., Mehler, B., Reimer, B., Lammers, V., D'Ambrosio, L. A., & Coughlin, J. F. (2010). The validity of driving simulation for assessing differences between in-vehicle informational interfaces: A comparison with field testing. *Ergonomics, 53*, 404–420.
Wierwille, W. W. (1993). Visual and manual demands of in-car controls and displays. In B. Peacock & W. Karwowski (Eds.), *Automotive ergonomics* (pp. 299–320). London, England: Taylor & Francis.
Wikman, A., & Summala, H. (2005). Aging and time-sharing in highway driving. *Optometry & Vision Science, 82*, 716–723.
Young, R. A. (2015). Need for revised total eyes-off-road criterion in the NHTSA distraction guidelines: Track radio-tuning data. In *Proceedings of the 8th international driving symposium on human factors in driver assessment, training, and vehicle design* (pp. 105–111). Iowa City, IA: University of Iowa.