Author(s): Kujala, Tuomo; Lasch, Annegret; Mäkelä, Jakke

Title: Critical Analysis on the NHTSA Acceptance Criteria for In-Vehicle Electronic Devices

Year: 2014

Version:

Critical Analysis on the NHTSA Acceptance Criteria for In-Vehicle Electronic Devices

Tuomo Kujala
University of Jyväskylä
Dept. of Computer Science and
Information Technology
P.O. Box 35, 40014 University of
Jyväskylä, Finland
+358400247392
tuomo.kujala@jyu.fi

Annegret Lasch
HERE Berlin
Invalidenstraße 116,
10115 Berlin, Germany
+4915165647169
annegret.lasch@here.com

Jakke Mäkelä
University of Jyväskylä
Dept. of Computer Science and
Information Technology
P.O. Box 35, 40014 University of
Jyväskylä, Finland
+358458408082
jakke.s.makela@jyu.fi

ABSTRACT

We tested a commercial in-car navigation system prototype against the NHTSA criteria for acceptance testing of in-vehicle electronic devices, in order to see what types of in-car tasks fail the acceptance test and why. In addition, we studied the visual demands of the driving scenario recommended by NHTSA for task acceptance testing. In the light of the results, NHTSA guidelines and acceptance criteria need to be further developed. In particular visual demands of the driving scenario and for different simulators need to be standardized in order to enable fair testing and comparable test results. We suggest the visual occlusion method for finding a driving scenario that corresponds better with real-life driving in visual demands as well as for standardizing the visual demands of the scenario when applied to different driving simulators. Furthermore, the acceptance criteria need to be re-evaluated. Especially the TEORT limit's applicability to a variety of test tasks needs to be validated and exceptions for certain task types considered. The utility of the average glance duration criterion should be reconsidered.

Categories and Subject Descriptors

H.5.2. Information interfaces and presentation: User Interfaces – Graphical user interfaces (GUI), Evaluation/methodology

General Terms

Measurement, Performance, Design, Experimentation, Human Factors

Keywords

Driver distraction, navigation system, touch screen, acceptance criteria, test performance, visual demand, visual occlusion.

1. INTRODUCTION

In 2013 National Highway Traffic Association (NHTSA) in US published the *Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices* [11] with the aim to decrease the amount of car crashes of drivers distracted by interacting with in-car electronic devices while driving. The guidelines consist of general interface recommendations and a collection of secondary task features that are considered to impede the safe operation of a car, and thus, should not be performable while driving. Further it outlines two test methods to determine the level of visual distraction induced by interactions with an in-car system, ultimately drawing a line on whether an in-car task is safe to be performed while driving.

The suggested methods for task acceptance testing are 1) *Occlusion Testing* or 2) *Eye Glance Measurement Using Driving Simulator Testing* [11]. The latter method, on which we are focusing here, uses metrics of eye glance durations away from the road while executing a task with the in-car system and driving in a simulator. Test participants are asked to follow a lead vehicle on a highway and maintain headway of around 70 m to it. The in-car test tasks are performed while on the move. The following criteria need to be fulfilled by an in-car task to pass the acceptance test ([11], p. 272 – 273):

1. *"For at least 21 of the 24 test participants, no more than 15 percent (rounded up) of the total number of eye glances away from the forward road scene have duration of greater than 2.0 seconds while performing the testable task one time.*

2. *For at least 21 of the 24 test participants, the mean duration of all eye glances away from the forward road scene is less than or equal to 2.0 seconds while performing the testable task one time.*

3. *For at least 21 of the 24 test participants, the sum of the durations of each individual participant's eye glances away from the forward road scene is less or equal to 12.0 seconds while performing the testable task one time."*

These criteria have been applied in research studies to assess distraction effects induced by in-car infotainment systems (i.e. [2][7]). However, so far there have been only a few analytical evaluations on how appropriate the NHTSA acceptance criteria

are as well as on the factors that influence whether a task passes the criteria or not. Furthermore, there is no solid theoretical background for the suggested driving scenario for task acceptance tests and no evidence that failure in the proposed testing criteria would in fact correspond to risky visual behaviors in this type of scenario. In other words, the visual demands of the suggested driving scenario are unknown. According to Wierwille's visual sampling model [17] drivers try to keep off-road glance durations between 500 to 1600 milliseconds in almost all driving situations in real traffic. There is no evidence to date indicating that this would hold also for the NHTSA driving scenario.

Visual occlusion is an established method for assessing the visual demands of driving [14][16]. The visual demand of driving can be defined as *the frequency at which the driver has to update the focus of visual attention in order to decrease uncertainty of the task-critical event states* (e.g. speed, lane position) *in the immediate field of view to a preferred level*, after [14] and [20]. In the visual occlusion technique the visual field of the driver (driving scene) is intermittently occluded by the means of visor, goggles or blanked screens on system- or driver-paced intervals in order to get an idea of the visual demands of driving. In driver-paced occlusion the occlusion time (OT) can be taken as equal to the longest time that a driver would choose to drive comfortably with eyes off road when fully concentrating on the driving task. 85th percentile OT of a driver sample can then be interpreted as a limit of acceptable behavior, following the logic of a common design standard in traffic engineering.

This paper focuses on the following research questions:

- Do the studied in-car navigation system functionalities pass the NHTSA recommended criteria?

- What are the reasons in-car tasks do not meet the criteria?

- What are the individually preferred limits of inappropriate off-road glance durations for the NHTSA recommended driving scenario as the 85th percentile occlusion times when concentrating fully on the driving task?

- Are these in line with the NHTSA recommended in-car glance duration distributions and acceptance criteria?

- Is the NHTSA driving scenario sensitive to small changes in the driving task?

In order to answer these questions, we evaluated a commercial navigation system based on the NHTSA criteria to see which in-car tasks would pass as well as studied the visual demands of the NHTSA driving scenario. In order to study the sensitivity of the NHTSA scenario's visual demands we compared these to a slightly modified scenario with no lead car.

Regarding the analyses of visual demands of the NHTSA scenario and the acceptance criteria 1 and 2, we tested the following hypotheses:

H1. The 85th percentile of the 85th percentile OTs is near 2.0 seconds (criterion 1).

H2. The 85th percentile of the ratio of over 2-second OTs is near 15 percent (criterion 1).

H3. The 85th percentile of the mean OTs is near 2.0 seconds (criterion 2).

H4. The OT distributions differ significantly between the NHTSA scenarios with and without the lead car.

2. METHOD
2.1 Design
The study followed the NHTSA (2013) task acceptance testing guidelines [11] as accurately as possible within the technical limitations caused by the driving simulator in use. The acceptance testing of the in-car tasks (10) followed a within-subject 10 x 1 design. For the analyses of the visual demands of the NHTSA scenario a within-subject 2 x 1 design was used (driving with and without a lead car).

2.2 Participants
A total of 26 participants took part in the study. Thirteen of the participants were female and the other half male. They were all in good general health. Participants were between 20 to 62 years old, with an average of 39.9 years (SD=15.2 years) and all of them had previous experience on cell phone use while driving. Six test participants were 18 to 24 years old, six 25 to 39 years old, six 40 to 54 years old, and eight were 55 years old or older. Two participants over 55 years felt simulator sickness and their data was removed from the test data. One had to cancel at the beginning of the tests whereas the other one was able to complete the occlusion trials before interrupting the test. All the participants had a valid driving license and were driving more than 5,000 kilometers (over 3,000 miles) per year. They were all unfamiliar with the prototype navigation system under testing. Participants were recruited by sending invitations through public university and company e-mail lists. Each participant was rewarded with a movie ticket for taking part in the study.

2.3 Apparatus
The tests were conducted at the Driving Simulator Laboratory of the Department of Computer Science and Information Systems in the University of Jyväskylä, Finland. The fixed-base medium-fidelity driving simulator consists of parts of a real vehicle cab (Figure 1).
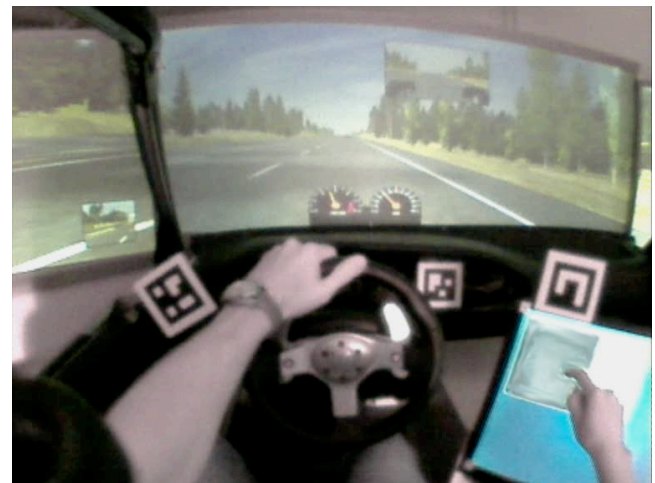
Figure 1. Experimental setup (navigation system UI blurred because of confidentiality issues)

The driving environment is projected on three screens. The front screen, positioned on a distance of 135 centimeters from the participants' eyes, has dimensions of 64 x 170 centimeters (1280 x 1024 pixels). From the NHTSA guidelines [11] (p. 205-206): "*As previously stated, the roadway display should be far enough in front of the simulator's driver that visual accommodation must occur when the driver switches his gaze between the device*

interface and the roadway. ... The minimum driver eye point to screen distance determined in this manner would be 1.27 meters. NHTSA has decided to round this 1.27 meter value up to 2.0 meters.* This means 1.35 meters should be acceptable even if the rounded minimum recommendation is 2.0 meters. The two side screens' dimensions are 64 x 110 centimeters (1280 x 1024 pixels). The left side screen is about 130 centimeters and the right side screen about 150 centimeters from the participant's eyes. The distances to the screens are fixed but the positions of the pedals and the steering wheel with force feedback are adjustable. The speedometer together with a RPM meter was projected on the driving scene above the dashboard. A large amount of telemetric data can be collected with the driving simulator software.

Driving scenario and environment complied with the NHTSA guidelines [11] as accurately as possible taking into account the limitations of the driving simulator. Driving simulator software (http://www.eepsoft.fi/) simulated driving on a long straight four-lane highway located in Southern Finland, with a divider to the opposite driving direction. There was no solid double yellow line down the center of the road but instead white dashed lane markings separating the lanes typical to Finnish highways. The driving simulator sent real-time simulated GPS data to the navigation system under testing to support guidance and search functionalities. Because of the GPS-simulation we needed to use a simulation of a real traffic environment. The part of the highway used for testing, E12 Hämeenlinnanväylä, is not perfectly straight but has a very mild curvature typical to a highway.

The prototype of the commercial navigation system software was displayed as a 10-inch head unit screen on the upper left corner of a 22–inch 3M capacitive multi-touch display. Dikablis head-mounted eye-tracking device with 50 Hz sampling rate was used for gathering eye movement data. Two Dell laptops were needed for capturing eye-tracking data and for controlling the experimental setup.

2.4 Procedure

Before the experiment started each participant had to sign a Non Disclosure Agreement and was informed about the purpose and setup of the study. They were then taken to the driving simulator and steering wheel and pedals were adjusted for the participant. Participants practiced driving in the simulator using a city environment. Once they felt comfortable driving the eye tracking system was put on and calibrated.

Participants were introduced to the navigation system and shown how to perform each of the tasks to be tested. They then could practice them until they felt comfortable doing them, without and while driving. An overview of the tested tasks is given in table 1. Tasks were grouped into 5 trials, with each trial consisting of 1 – 3 tasks.

Table 1. In-car tasks under testing

Task (N)	Instructions and task steps
1. Find road number (23)	Find and read aloud the number of the road you are driving on.
2. Get info on current location (23)	Get information on your current location. 1. Tap on current location indicator 2. Tap on back button
3.1. Follow guidance (23)	NB: Before participants started driving they were asked to set the destination to "Klaukkalantie 10". Follow the guidance instructions. 1. View visual guidance indicator on map
3.2. Repeat voice guidance (23)	Repeat the last voice guidance you've heard. 1. Tap on visual guidance indicator on the map
3.3. Mute guidance (23)	Mute guidance. 1. Select settings menu 2. Deselect voice guidance 3. Tap on back button
3.4. Stop guidance (23)	Stop guidance. 1. Select settings menu 2. Tap on stop button 3. Confirm that guidance should be stopped
4.1. Add hotel to your route (22)	NB: Before participants started driving they were asked to set the destination to "Klaukkalantie 10" and add the shop "Suomussalmen Kiriakauppa" as a waypoint to the route. Add the hotel "Easydays/Lomamökit" as a waypoint to the current route and start guidance. 1. Select Search 2. Select Hotels from POI list 3. Select target hotel 4. Tap on Add waypoint button 5. Tap on Start guidance
4.2. Reorder waypoints (22)	Reorder the waypoints and start guidance. 1. Select settings menu 2. Tap on Edit route 3. Tap on up arrow next to waypoint to move it up
4.3 Delete the hotel waypoint (22)	Delete the waypoint "Easydays/Lomamökit" and start guidance. 1. Tap on delete button next to waypoint 2. Tap on Done
5. Get guidance to a restaurant without using toll roads (18)	Start guidance to a restaurant "Keimola Golf" via roads without tolls. 1. Select Search 2. Select Eat&Drink from POI list 3. Select target restaurant 4. Tap on Set destination 5. Tap on Route options 6. Uncheck toll roads 7. Tap on back button 8. Tap on Start navigation

Furthermore, the participants received instructions on the driving task, which should be prioritized over the secondary task. They were asked to drive in the right hand lane and keep a constant headway distance of 70 meters to the lead vehicle driving at a speed of 80 km/h (50 mph). Participants then could practice driving the highway scenario.

Order of the trials were mixed and counterbalanced. Each trial took about 3.5 minutes each. Testing of a task was not done for a participant if the task had already failed a criterion or there were enough successful performances (21/24) to pass all the criteria.

After the NHTSA task acceptance trials participants went on with two visual occlusion trials, driving occluded in the NHTSA scenario as well as in a modified NHTSA scenario without the lead car. The order of the trials were counterbalanced across the

sample. In the occlusion trial, the screens were occluded (black) as default, and the participant could unocclude the driving scene for 500 ms for each press of the right side paddle shifter in the steering wheel. The goals for the two trials were to keep the own lane and a constant 70-meter headway distance to a lead car driving at a constant 80 km/h (50 mph) for the NHTSA scenario or a constant speed of 80 km/h (50 mph) for the scenario without the lead car.

2.5. Analysis

Due to tracking inaccuracies and the pupil often being lost during eye-movements from the driving scene to in-car display and back, the automatically scored glance durations seemed to be systematically less than manually scored durations. Thus, a human data reducer determined from the overlaid gaze video data the off-road glance durations by following the SAE-J2396 standard [15].

Occlusion time (OT), which is the time that the driver drove with vision occluded, was used to estimate the visual demand. The data was calculated with a script from the driving simulator log data. It is clear that a longer occlusion time corresponds to a lower perceived visual demand, even though the exact functional form is not known. Even without knowing the exact form, it is possible to compare the OT distributions of the two scenarios to determine which scenario has a higher visual demand. OT distributions are typically heavy-tailed, and thus Gaussian statistics cannot be properly used. In these cases statistical analyses were performed with a Wilcoxon rank sum test with continuity correction for the median OTs and Welch two-sample t-test on the logarithm of the OT. Alpha level of .01 was used in the statistical testing.

3. RESULTS

3.1 Acceptance test results for the in-car navigation system

Mean individual glance duration, mean percentage of over 2 seconds glances and mean Total Eyes Off Road Time (TEORT) are shown in Table 2. There were no off road glances to elsewhere than in the head unit, so in-car glances equal eyes off road glances here (e.g., no glances on mirrors).

Table 2: Mean individual glance duration, Mean percentage of over-2-second in-car glances, Mean TEORT in seconds

| Task (*N*)	Mean individual glance durations (SD)	Mean percentage of >2s in-car glances (SD)	Mean Total Eyes Off Road Time in seconds (SD)
Find road number (23)	.92 (.29)	1.40 (5.36)	8.06 (5.38)
Get info on current location (23)	.77 (.24)	0.00 (0.00)	2.72 (1.00)
Follow guidance (23)	.53 (.32)	0.00 (0.00)	1.68 (2.19)
Repeat guidance (23)	.75 (.39)	0.00 (0.00)	1.01 (0.54)
Mute guidance (23)	.90 (.32)	1.09 (5.21)	2.99 (1.26)
Stop guidance (23)	.66 (.26)	0.00 (0.00)	2.36 (1.34)
Add a hotel to your route (22)	.89 (.24)	1.06 (2.78)	15.64 (7.72)
Reorder waypoints (22)	.81 (.29)	2.85 (6.68)	5.61 (2.14)
Delete the hotel waypoint (22)	.81 (.24)	0.00 (0.00)	6.21 (2.28)
Get guidance to a restaurant without using toll roads (18)	.81 (.27)	1.70 (5.69)	12.13 (3.71)

Table 3 gives an overview of the number of participants failing the criteria. The mean individual glance duration criterion was passed for all the tasks. The percentage of over 2 seconds glances was above the 15% limit for 1 out of 24 participants for 4 tasks. More than 3 out of 24 participants exceeded the 12 seconds TEORT limit for the tasks "*Find road number*" (5), "*Add a hotel to your route*" (14) and "*Get guidance to a restaurant without using toll roads*" (8), thus failing the acceptance test.

Table 3. NHTSA test results per in-car task, number of participants failing the criterion (failed tasks in bold)

Task (*N*)	Mean of individual glance durations more than 2.0 seconds	85% of individual glance durations less than 2.0 seconds	Sum of individual eye glance durations less than or equal to 12.0 seconds
Find road number (23)	0	1 (4.3%)	**5 (21.7%)**
Get info on current location (23)	0	0	0
Follow guidance (23)	0	0	0
Repeat guidance (23)	0	0	0
Mute guidance (23)	0	1 (4.3%)	0
Stop guidance (23)	0	0	0
Add a hotel to your route (22)	0	0	**14 (63.6%)**
Reorder waypoints (22)	0	1 (4.5%)	0
Delete the hotel waypoint (22)	0	0	1 (4.5%)
Get guidance to a restaurant without using toll roads (18)	0	1 (5.6%)	**8 (44.4%)**

3.2 Visual Demands of the NHTSA Scenario

As expected, the OT distributions were found to be heavy-tailed (Figure 2), and thus Gaussian statistics cannot be properly used. Robust estimators (median, 85th percentile) can, however, be determined (Table 5).
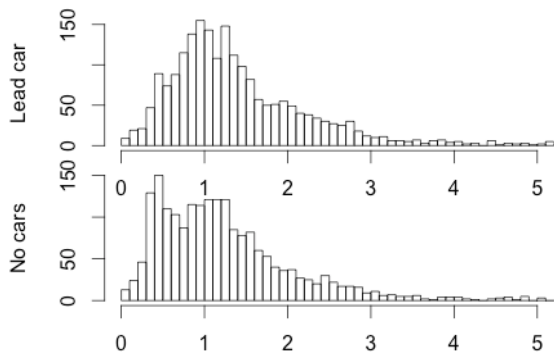
Figure 2. OT distributions in the NHTSA scenario (up, *N*=2069) and in the modified NHTSA scenario without the lead car (down, *N*=1992), s

Table 5. Occlusion descriptives, median, mean (sd), *N*=25

Driving Scenario	OT, s	Speed, m/s	85th% of 85th% OT, s	85th% of over 2 s OT, %	85th% mean OT, s
NHTSA (2013)	1.22 1.42 (0.90)	22.0 22.1 (1.3)	2.54	46	2.00
NHTSA (2013) no lead car	1.08 1.26 (0.89)	22.2 22.2 (1.1)	2.14	38	1.87

Considering H1, the data indicates the 85th percentile of the 85th percentile OTs in the NHTSA scenario is 2.54 seconds which means just over 500 millisecond difference to the bar of 2.0 seconds (Table 5 and Figure 3). The mean difference is small but the difference in the 85th percentile ratios of over 2 seconds OTs in the NHTSA scenario is unacceptably high (46%) compared to the bar of 15 percent (H2, Table 5, Figure 4). There is one participant with a value looking as an outlier, but there is no clear reason to reject him. There are several others with over 15% values as can be seen in Figure 4.
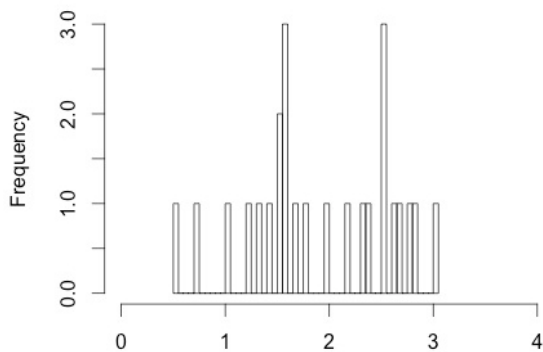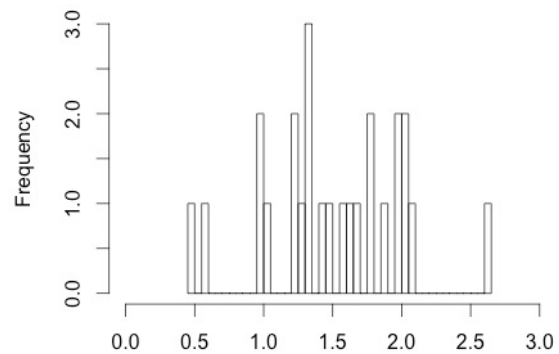

**Figure 3. Individual 85th percentile OTs in the NHTSA scenario, s**
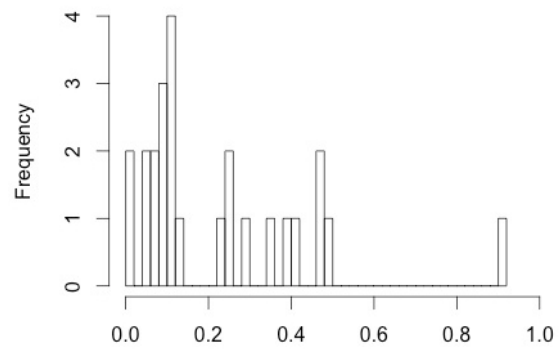
Figure 4. Individual over 2-second OT ratios in the NHTSA scenario

H3 gets direct support from the occlusion data. The 85th percentile of the mean OTs for the NHTSA scenario was exactly 2.0 seconds (see Figure 5).

Figure 5. Individual mean OTs in the NHTSA scenario, s

Regarding H4, a Wilcoxon rank sum test with continuity correction suggests that the medians of the two driving scenarios are significantly different (*W*=1770226, *p*<.001). More information can be found by taking the logarithm of the OT values. This distribution was found to be symmetric though non-Gaussian when the shortest outliers were removed (Figure 3.). Statistical fits can thus be performed on the logarithm of the OT. A Welch two-sample *t*-test shows a significant difference in the mean values of mu* between the driving scenarios: $t(3950.821)=-0.78662$, *p*<.001, 95% CI [-0.201, -0.121]. OTs were somewhat higher in the case of the original NHTSA scenario with a lead car, indicating lower visual demands.
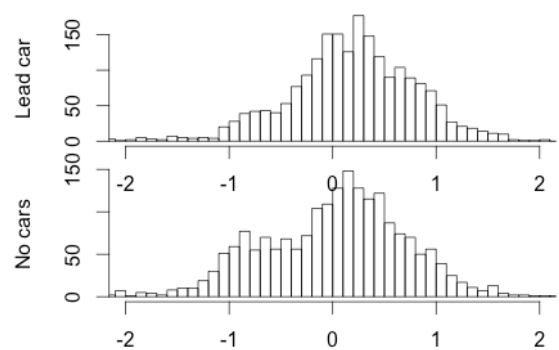
Figure 6. Logarithmic OT distributions in the NHTSA scenario (up) and in the modified NHTSA scenario without the lead car (down)

In km/h, for the NHTSA scenario with the lead car the mean speed was 79.6 (SD=4.7), whereas for the scenario with no lead car it was 79.9 (SD=4.0). The car speeds are symmetrically distributed, and Gaussian statistics can be used directly (Figure 4). The means differ in a statistically significant way, but in practice the effect is too small to be meaningful: a two-sample Welch test gives $t(4005.216)=4.399$, $p<.001$, 95% CI [0.091, 0.24]. An F-test for variances, however, gives a statistically significant ratio of variances of 1.36 with $F(1991,2068)=0.734$, $p<.001$, 95% CI [1.25,1.49]. The speeds varied more when there was a lead car.
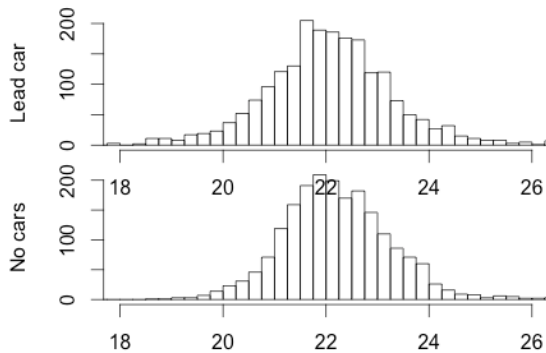
Figure 7. Vehicle speed distributions in the NHTSA scenario (up) and in the modified NHTSA scenario without the lead car (down), m/s (22.22 m/s ~ 80 km/h ~ 50 mph)

The higher variance of speed with the lead car could be a result of task instructions. In the NHTSA scenario the goal is to keep a steady 70 m distance to the lead car driving at 80 km/h, whereas in our modified scenario without the lead car the goal was to keep a steady speed of 80 km/h (in addition to lane-keeping in both scenarios). For the distance keeping task the goal of 70 m headway is more vague and thus, probably more uncertainty could be tolerated in the task than in the speed maintenance task when the speedometer needle told the driver the accurate speed value. The greater visual demands of the scenario without the lead car could be explained by the requirement to shift gaze between the vanishing point of the road for anticipatory look-ahead fixations [8] and the speedometer necessary for speed control in the fixed-base simulator. The anticipatory information gathering and speed control (in the form of headway distance control) can be visually more easily managed when the driver is able to keep the fixations on the lead car moving at a static speed. The uncertainty about the states of speed and the path of travel can be reduced simultaneously when keeping the fixations on the tail of a lead car.

4. DISCUSSION

In this experiment, we evaluated ten in-car tasks on a commercial in-car navigation system prototype following the NHTSA task acceptance testing of eye glance measurements using a driving simulator [11], in order to see what types of in-car tasks fail the test and why. We further studied the visual demands of the driving scenario recommended by NHTSA for acceptance testing.

Out of the ten tested tasks seven easily passed the acceptance criteria, but three of them exceeded the Total Eyes Off Road Time (TEORT) limit of 12 seconds. This included the tasks "*Add a hotel to your route*" and "*Get guidance to a restaurant without using toll roads*" which indeed require a high number of interaction steps. It is recommended to reduce the amount of steps needed to perform these tasks to reduce the TEORT. Surprisingly

also the task "*Find road number*" did not pass although it did not require any manual interaction with the system, instead the driver only needed to identify the road number from the map. However, the road number was not continuously shown on the map, thus it was a matter of chance if a participant happened to look at the display at the very moment the road number was present. The test result could imply to present this information continuously. However, this would result in visually cluttered maps, considering that there is other relevant information in addition to the road number that would need to be presented permanently. Instead it seems that for certain types of in-car tasks the TEORT limit just seems not to work very well.

Our test tasks 4.1 and 5 had the search targets listed early in the search results. However, there are search tasks, such as looking for a nearby restaurant or a gas station that can provide a long list of search results. It matters a lot from the TEORT perspective where the target is located in the result list. If the search target is put on the first page of the list, the 12 seconds TEORT criterion could easily be passed. However, how likely is it that the target is always on the first page? Does this mean we should test all possible target positions against TEORT? In a similar manner there are real-world tasks for which the end of a task cannot be defined exactly, i.e. exploring the content of a music player. Already today one can easily browse internet radio stations for more than 12 or 20 seconds just to find one you like among the numerous possibilities available. These types of in-car services that are available in many infotainment systems today, need to be tested for distraction as well, and should have user interfaces optimized to minimize the associated risky overlong glances.

Given these limitations it can be argued that it needs to be clearly defined for which types of in-car tasks TEORT is applicable as it obviously is not suitable for any given task. If applied to the example browsing tasks above, following TEORT recommendations would ultimately mean to hide away a lot of information, e.g., about services nearby, which are however much valued by drivers. In addition, it is unclear how well TEORT applies for multi-step and extended interactions that characterize activities involving voice-command interfaces with visual feedback elements [13]. The good point in TEORT for search tasks is that it forces designers to optimize the search interfaces (short cuts, sorting, etc.) so that the search target won't be at the bottom of the list, unless there is no target (i.e., in browsing music).

From the risk association and visual demand perspective, the percentage of over 2 seconds glance criteria seems more valuable than the TEORT limit, and should be decisive for these kinds of tasks. Yet, this criterion is only meaningful if the visual demands of the driving scenario are adjusted to correspond to the real visual demands of driving. This is why we analyzed the visual demands of the NHTSA driving scenario by testing four hypotheses.

Regarding hypotheses H1 and H2, our analyses of the visual demands of the NHTSA scenario (at least as realized in our simulator) indicated that the driving task seems to be very low on its visual demands. 85 percent of the participants felt comfortable driving blind for more than 2 seconds at a time in about 46% of the occluded periods. Compared to earlier studies on visual demands of driving [14][17], the NHTSA scenario does not really well imitate the visual demands of real driving. Well, who would try to keep a constant 70-meter distance to a lead car while keeping the own lane in real life without observing anything else in the environment? The 85th percentile for the 85th percentile

occlusion times was as high as 2.54 seconds. Does this mean the acceptance limit should be 2.5 seconds instead of 2 seconds in this driving scenario?

From the safety point of view it is of course rational to keep the criterion somewhat under the driver-accepted OTs, but here the 85th percentiles of over 2 seconds OT ratios were certainly too high compared to real driving. One reason for the low visual demands could be that there is no risk of collision or unexpected events as in more realistic driving scenarios. Keeping the gaze at the lead car is all that is required in order to keep the distance and the lane position for experienced drivers [18] who are to be recruited by the NHTSA guidelines.

The data gives direct support for H3. The 85th percentile of the mean OTs for the NHTSA scenario was exactly 2.0 seconds. However, when looking at our test results, the results of [4], and the Wierwille's visual sampling model for real driving [17], this acceptance criteria does not make much sense. The average in-car glance durations for the test tasks were all below 1.0 seconds and in line with Wierwille's model. One might wonder what kind of horrible in-car task would *not* pass the limit, and thus it can be considered as useless. The fact that the criterion was supported by the visual demand data actually tells us again that the visual demands of the NHTSA driving scenario are too low.

According to literature, there is a rationale behind the NHTSA 2 seconds limit for a risky off road glance. Field studies have indicated a statistical relationship between individual 2-second off-road glance durations and the risk of safety-critical incidents [9]. According to Wierwille [17] drivers try to keep in-car glance durations between 500 to 1600 milliseconds in almost all driving situations. But shouldn't we make sure that the driving scenario in any given driving simulator used for task acceptance testing corresponds to a real-world scenario, in which 2 seconds is near the 85th percentile for OT?

H4 got support from the data. The modified NHTSA scenario without the lead car seemed to have somewhat greater visual demands. The demands of this scenario were closer to the criterion of NHTSA (85th % of over 2 s OT: 38%, 85th % of 85th % OTs: 2.14 s). One reason for the lower visual demands of the NHTSA scenario might be that keeping the gaze at the lead car is all that is required in order to control the headway distance and the lane position. Without the lead car the driver has to alternate fixations between the road ahead for anticipatory information gathering [8] and the speedometer for speed control in a fixed-base simulator. However, the visual demands of the scenario without the lead car would be much more sensitive to simulator-dependent factors such as the location of the speedometer as well as motion feedback of the simulator. With a motion platform, the need to observe the speedometer for keeping a static speed can be significantly lower than in a fixed-base simulator. A question for further research is: would the visual demands have been reversed if the drivers were simply told that the speed limit in the simulated highway is 80 km/h but with no explicit instructions to monitor and attempt to maintain accurately the speed?

Based on these findings, one can raise the question whether the visual demands of the NHTSA driving scenario and the test results would be comparable if the tests were run in different simulators? There were small differences in our driving scenario compared to the NHTSA scenario, most importantly the lane markings were white, not yellow, and there were small curvature on the road instead of a perfectly straight road. However, it seems unlikely that these differences could have a major impact to the visual demands of the driving scenario. In fact these differences should raise the demands if anything, not lower. In any case, we can highly recommend the use of the visual occlusion method for controlling the visual demands of the NHTSA testing scenario across different simulator platforms.

When looking at the visual demand data carefully one can notice significant individual differences in the experienced levels of visual demand. These could be due to personal factors such as uncertainty toleration [4] or the skill level of the driver [18]. The usage of the often-used 85th percentile in the acceptance criteria shows a pretty nice sense of realism about the data. Medians and percentiles are very robust estimators whenever the data is heavy-tailed, very messy and sample sizes are small. And that is what we have here. Again the use of the mean in-car glance duration as an acceptance criterion can be highly questioned.

NHTSA guidelines are not perfect but they are the best we have at the moment for controlling the explosion of bad, even lethal, in-car user interfaces. Without this kind of guidance, even if imperfect, we could really talk about "killer apps". In the current study, the criteria were able to differentiate the most visually demanding tasks from the tasks with low visual demands (with one false exception). Thus the testing procedure and criteria allow in-car system designers to iteratively test and validate their design to ultimately create less demanding interfaces and interactions. Changes to the UI have been already made based on the results.

Given the visual demands of the driving scenario in our simulator, the criterion was strict regarding the percentage of long in-car glances, which was a good thing. Interestingly however, all the in-car tasks passed this criterion although there were individual percentages exceeding 15 percent. In some simulator implementations, however, it could be that the visual demands of the NHTSA driving scenario are much higher. In the worst-case scenario this would mean that visually highly demanding in-car tasks could pass the percentage of over 2 seconds glance criterion due to unrealistically high visual demands of driving. At a general level, drivers try to adapt the in-car glance durations according to the prevailing visual demands of the driving task [17].

AAM [1] as well as NHTSA [11] have used a manual radio tuning task as the baseline reference task for determining a societally acceptable level of distraction. This type of task could help to calibrate the test results for different simulators. However, NHTSA does not specify exactly what radio and what kind of task should be used in the baseline tests. Even if the radio and the baseline reference task would be accurately specified, this won't solve the problem of possibly varying visual demands of the driving scenario in different driving simulators. The radio tuning task won't give reliable baseline to the visual demands of the driving scenario against which inappropriate off-road glance durations could be decided. A more theoretically sound base for the criteria is needed. Driver's self-accepted 85th percentile OTs when fully focusing on driving could be used as a limit for acceptable off-road glancing behavior. In-car glance durations exceeding the self-accepted OTs would then indicate a real lapse of control. This approach would give room or means to account for individual differences between test samples. Significant differences in the test participants' visual sampling skills, especially with elderly drivers sample, are highly probable [19].

In our study we also noticed that many eye-tracking systems give unreliable automatic scores for the off-road glance durations when Area-Of-Interest analyses are used. Due to inaccuracy and because the pupil is often lost during eye-movements from the

driving scene to in-car display and back, the glance durations seemed to be systematically less than the manually scored durations. Despite the increased work effort we suggest to use manual scoring of the off-road glances following the SAE-J2396 [15] definitions in order to get reliable and comparable glance durations (see also [13]). The automatically scored glance duration data should in any case be carefully inspected manually. As an important end note, the visual occlusion acceptance test alternative outlined in the NHTSA guidelines has little to do with the visual occlusion method suggested here, and can be highly criticized for its metrics' informativeness and validity [6][10][12].

5. CONCLUSION

In light of the findings the NHTSA guidelines and acceptance criteria should be further developed. In particular visual demands of the driving scenario and for different simulators need to be standardized in order to enable fair testing and comparable test results. We suggest the visual occlusion method [14] for finding a driving scenario that corresponds better with real-life driving in visual demands as well as standardizing the visual demands of the scenario when applied to different driving simulators. In addition, alternative, less expensive scenarios with less irrelevant details could be used in acceptance testing as long as the visual demands correspond to those of the NHTSA scenario.

Furthermore, the acceptance criteria need to be re-evaluated taking into account all the possible modern in-car activities. Especially the TEORT limit's applicability to a variety of test tasks needs to be validated and exceptions for certain task types considered. The utility of the average glance duration criteria should be reconsidered. Specifications on test task design for acceptance testing would be helpful to avoid manipulation of test results into a desired direction by task design.

ACKNOWLEDGMENTS

REFERENCES

[1] Alliance of Automobile Manufacturers (2006). *Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems.*

[2] Broström, R., Aust, M.L., Wahlberg, L., & Källgren, L. (2013). What drives off-road glance durations during multitasking – capacity, practice or strategy? In *Proceedings of the 3rd International Conference on Driver Distraction and Inattention* (September 4-6, 2013, Gothenburg, Sweden).

[3] Dingus, T.A., Klauer, S.G., Neale, V.L., Petersen, A., Lee, S.E., Sudweeks, J., et al. (2006). *The 100-Car Naturalistic Driving Study. Phase II – Results of the 100-Car Field Experiment.* (DOT HS 810 593). Washington DC: NHTSA.

[4] Dopart, C., Häggman, A., Thornberry, C., Mehler, B., Dobres, J., and Reimer, B. (2013). A driving simulation study examining destination entry with iPhone iOS 5 Google Maps and a Garmin portable GPS system. In *Proceedings of the 57th Annual Meeting of the Human Factors and Ergonomics Society*, San Diego, CA, Sept. 30 - Oct. 4, 2013.

[5] Furnham, A. & Marks, J. (2013). Tolerance of Ambiguity: A Review of the Recent Literature. *Psychology, 4*, 717-728.

[6] Kujala, T. (2009). Occlusion technique – Valid metrics for testing visual distraction? In *Proceedings of the European Conference on Cognitive Ergonomics 2009* (pp. 341-348). VTT Symposium series 258, VTT Technical Research Centre of Finland.

[7] Kujala, T., Silvennoinen, J., Lasch, A. (2013). Visual-manual in-car tasks decomposed: text entry and kinetic scrolling as the main sources of visual distraction. *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Eindhoven, Netherlands, 28. – 30.10.2013), 82-89.

[8] Lehtonen, E, Lappi, O, Kotkanen, H, & Summala, H. (2013). Look-ahead fixations in curve driving. *Ergonomics, 56*, 34-44.

[9] 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, 4*, 1104-1116.

[10] Monk, C.A. & Kidd D.G. (2007). R we fooling ourselves: Does the occlusion technique shortchange R estimates? In *Proceedings of the Fourth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (Stevenson, WA, July 9-12, 2007), 2-8.

[11] National Highway Traffic Safety Administration (2013). *Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices.* (NHTSA-2010-0053).

[12] Noy, I., Lemoine, L., Klachan, C., and Burns, P.C. (2004). Task interruptability and duration as measures of visual distraction. *Appl. Ergon. 35*, 207-213.

[13] Reimer, B., Mehler, B., Dobres, J. and 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. Massachusetts Institute of Technology, Cambridge, MA.

[14] Senders, J. W., Kristofferson, A. B., Levison, W.H., Dietrich, C. W., & Ward, J. L. (1967). *The attentional demand of automobile driving. Highway Res. Record, No. 195*, 15-33.

[15] Society of Automotive Engineers (2000). *SAE-J2396 Definitions and Experimental Measures Related to the Specification of Driver Visual Behavior Using Video Based Techniques.* Warrendale, PA: SAE.

[16] Tsimhoni, O. and Green, P. (1999). Visual demand of driving curves determined by visual occlusion. *Proceedings of the Vision in Vehicles 8 Conference* (Boston, MA, August 22-25).

[17] Wierwille, W. W. (1993). An initial model of visual sampling of in-car displays and controls. In A.G. Gale, I.D. Brown, C.M. Haslegrave, H.W. Kruysse, & S.P. Taylor (Eds.), *Vision in vehicles IV* (pp. 271-279). Elsevier.

[18] Wikman, A.S., Nieminen, T. & Summala, H. (1998). Driving experience and time-sharing during in-car tasks on roads of different width. *Ergonomics, 41*, 358-372.

[19] Wikman, A. S. & Summala, H. (2005). Aging and time-sharing in highway driving. *Optometry Vis. Sci. 82*, 716-723.

[20] Wortelen, B., Baumann, M., & Lüdtke, A. (2013). Dynamic simulation and prediction of drivers' attention distribution. *Transportation Research Part F, 21*, 278-294.