Eyes on the Road!

Off-Road Glance Durations when Performing Tasks on In-Vehicle Systems while Driving in a Simulator

Linnea Wahlberg
2013-06-14

ISRN-number: LIU-ID/A-KOGVET-A--13/007--SE
Supervisor: Björn Peters
Co-supervisor: Mikael Ljung Aust, Volvo Cars
Examiner: Arne Jönsson
Reviewer: Georg Abadir Guirgis
Abstract
The 85th percentile off-road glances while performing three tasks on an in-vehicle system while driving in a simulator was investigated. The tasks were a radio task, a telephone task and a sound settings task which were performed at three occasions each. The distribution of 85th percentile off-road glance durations for each subject and task showed that durations differed between individuals rather than between tasks. It also turned out that durations longer than 2.00 seconds were not rare and 2 of 16 subjects had durations longer than 2.00 seconds in the radio task. Even though the distribution showed small differences between tasks on an individual level, differences on a group level were found between the tasks. A tendency of a learning effect was found, which implied a decrease in 85th percentile off-road glance durations as the tasks were performed at several occasions. A tendency of a floor effect in 85th percentile off-road glance durations, when the subjects are familiarized with tasks, was also found. Performance on a computerized trail-making test, measuring ability of visual search, motor speed and mental flexibility, was found not to be related with 85th percentile off-road glance durations.
Acknowledgement
I would like to thank Björn Peters for his great support and help throughout my thesis work, and Mikael Ljung Aust for his help with the design and execution of the study, as well as for input and feedback during the writing process. I would like to thank Laban Källgren for his programming and helping me filter the data and Anne Bolling for helping me find participants and for assisting me during the experiments. I would also like to thank Heikki Summala for his input regarding the trail-making test. Finally I would like to thank the participants for taking their time to make this study possible.
Contents

1. INTRODUCTION ............................................................................................................................................. 1

2. PURPOSE OF THE STUDY .............................................................................................................................. 7

3. METHOD ........................................................................................................................................................... 9

   3.1 EXPERIMENTAL DESIGN .......................................................................................................................... 9
   3.2 SUBJECTS .................................................................................................................................................... 9
   3.3 APPARATUS ................................................................................................................................................. 9
   3.4 DRIVING TASK AND SECONDARY TASKS ............................................................................................... 11
   3.5 TRAIL-MAKING TEST .............................................................................................................................. 14
   3.6 PROCEDURE .............................................................................................................................................. 15
   3.7 MEASURES ............................................................................................................................................... 16
   3.8 DATA FILTERING .................................................................................................................................. 16
   3.9 STATISTICAL ANALYSIS ......................................................................................................................... 18

4. RESULTS ......................................................................................................................................................... 19

   4.1 VARIATION IN 85TH PERCENTILE OFF-ROAD GLANCE DURATIONS .................................................. 19
   4.2 DIFFERENCES BETWEEN TASKS AND OCCASIONS .......................................................................... 20
   4.3 INDIVIDUAL DIFFERENCES AND 85TH PERCENTILE OFF-ROAD GLANCE DURATIONS .................. 22
   4.4 GLANCES AT THE FORWARD ROAD ..................................................................................................... 24
   4.5 QUESTIONNAIRE ................................................................................................................................... 24

5. DISCUSSION ................................................................................................................................................... 25

   5.1 VARIATION IN 85TH PERCENTILE OFF-ROAD GLANCE DURATIONS .................................................... 25
   5.2 DIFFERENCES BETWEEN OCCASIONS ................................................................................................. 25
   5.3 DIFFERENCES BETWEEN TASKS .......................................................................................................... 26
   5.4 TRAIL-MAKING TEST AND 85TH PERCENTILE OFF-ROAD GLANCE DURATIONS ............................ 27
   5.5 AGE AND 85TH PERCENTILE OFF-ROAD GLANCE DURATIONS ....................................................... 28
   5.6 LIMITATIONS ....................................................................................................................................... 28

6. CONCLUSION .................................................................................................................................................. 31

REFERENCES .................................................................................................................................................. 33

APPENDIX A: DATA FILTERING .................................................................................................................. 35
1. Introduction

Drivers are exposed to an increasing number of technical systems in cars today. There are technical systems with the purpose of simplifying the driving task, with the purpose of improving the driving experience and even improving the quality of the drivers’ lives. These systems are called in-vehicle infotainment systems. Examples include systems that help the driver, for example with navigation, and that entertain the driver, for example by playing music. There is however a drawback with these systems, as they often require attention and actions from the driver, so that the driver has to divide his/her attention between operating the infotainment system and operating the vehicle. This leads to reduced attentional resources for each of the two tasks compared to performing each task alone (Metz, Schömig, Krüger, 2011).

An especially important part of attention in driving is visual attention. Drivers get much of their understanding of the situation from looking at the traffic environment, which include elements such as the road and other vehicles. The driver updates his/her current understanding of the situation based on this visual information. S/he can then decide what actions to take to achieve his/her goals, based on the understanding of the situation. The actions change the situation, i.e. the system status, and therefore the driver needs to update his/her current understanding again.

This loop is in accordance with Hollnagel’s (1999) contextual control model, COCOM. The driver’s visual input, and interpretation of the input, gives the driver information that helps him or her to keep control over the vehicle and the situation. When the visual scanning resources for the driving task decreases because of a visual secondary task, the driver gets less information to build his/her understanding on, which can lead to inappropriate actions resulting in unsafe driving behavior. COCOM does not model cognition and underlying processes, but rather focus on the observed variety of the operator to analyze controller performance. The model does however include the construct competence, which is a small range of functions or capabilities of the controller. When the controller is human this corresponds to skills, templates and cognitive functions, such as observation, interpretation, planning and execution. COCOM does however not consider individual differences between human controllers. (Hollnagel, 1999)

Even though it is important to include situational factors in models of driving behavior, Ranney (1994) considered it very unlikely that crash causation only involves transient, situation-specific, factors. He reported that theories that accounted for the inherent variability of human behavior may be more important to roadway crash causation than systematic errors and mentioned that individual traits such as perceptual style and selective attention have been found to predict accident involvement. Visual attention, measured with the useful field of view, specifically has been found to predict accidents in older drivers (Ball, Owsley, Sloane, Roenker & Bruni, 1993; Owsley, Ball, Sloane, Roenker, Bruni, 1991).

The ITERATE model (Oppenheim et al., 2012) is a driver-centered model of driver behavior, which is based on the idea that situational differences and intra and inter individual differences affect driver behavior. The model includes personality, experience, driver state,
task demand and culture as driver variables and considers environmental and vehicular parameters as inputs to the driver. The driver variables vary between individuals and situation and were selected based on previous research that had been found affecting driver behavior.

Ability of visual search, motor speed and mental flexibility are other individual factors that are likely important in driving, and may be especially important when the driver is performing in-vehicle system tasks. These factors (Crowe, 1998), and related factors such as the cognitive domains of processing speed, sequencing and visual-motor skills (Bowie & Harvey, 2006), are considered to be measured with a trail-making test. Differences in performance in a trail-making test could therefore explain some individual differences in visual behavior while driving.

In the traditional form, the trail-making test consists of two parts, where the variable of interest is how long the subject takes to perform each part. In part A, the subject connects a series of encircled numbers in a numerical order, by drawing a line between them. In part B, the subject draws a line to connect a series of encircled numbers and letters in a combined numerical and alphabetical order, alternating between numbers and letters according to the sequence 1-A-2-B-3-C and so on. (Bowie & Harvey, 2006) Among nonclinical individuals, part A has been found to measure visual search and motor speed, and part B has been found to measure visual search and mental flexibility (Crowe, 1998). The trail-making test has been used to predict driving ability among older drivers (Emerson, et al. 2012) and individuals after brain injury (Hargrave, Nupp & Erickson, 2012).

A computerized version of the trail-making test (Summala, Etholén, Leino, Niskakangas, Laine & Saarinen, 2008) have been developed, which is performed on a touch screen, on which the subject taps on the targets instead of drawing a line between them. In this test there are two versions of part A and part B. One version entails targets with fixed positions and one version entails targets that randomly change positions after the subject has tapped on a target. Hence, the computerized trail-making test consists of four sub-tests: part A fixed, part A random, part B fixed and part B random.

The versions with randomized targets should demand more mental flexibility and therefore be more difficult than the versions with fixed targets. By subtracting mean time between correct responses in part A fixed from the mean time between correct responses in part B fixed, the ability of mental flexibility controlled for the visual search component can be measured. By subtracting mean time between correct responses in part A fixed from mean time between responses in part A random, the gain from having fixed targets, i.e. how much one benefits from stable targets in relation to random targets, can be measured. This would result in a more refined measure. Visual search, motor speed and mental flexibility are important in driving. It should therefore be investigated whether performance in the computerized trail-making test is related with visual behavior while performing a visual and motoric task while driving.

When performing a visual and motoric task while driving the driver is distracted and shifts his/her attention from the traffic situation to the system, on which the task is performed. Studies have shown that drivers who are distracted longer than 2.00 seconds drive less safely than when they are not distracted. Distractions with durations less than 2.00 seconds did
however not affect safe driving (Ryu, Sihn, Yu, 2013). Using data from the 100-Car Naturalistic Driving Study (Dingus et al., 2006), it has been found that drivers who look away from the road longer than a total of 2.00 seconds, in a period of 5.00 seconds before and 1.00 second after a precipitating event, more than double the risk of crashes and near-crashes (Klauer, Dingus, Neale, Sudweeks, Ramsey, 2006). Moreover, a later study (Liang, Lee, Yekhshatyan, 2012), which used the same dataset, has found that the most sensitive estimation of risk is based on lengths of instantaneous off-road glance durations, i.e. lengths of off-road glances during the precipitating event rather than during a time period around the event. Glance history did not increase prediction precision. These results imply that how long the driver looks away from the road while performing tasks on in-vehicle systems is very important to consider when developing in-vehicle systems. In particular, interactions that require off-road glances longer than 2.00 seconds are especially critical to avoid.

Because of the risks that accompany looking away from the road while driving, the American National Highway Traffic Safety Administration, NHTSA, (2012) has proposed three acceptance criteria for testing visual-manual interaction with non-driving related in-vehicle systems. To test a system, a testable secondary task, i.e. a task with a definable start and ending, is performed on the in-vehicle system while driving in a simulator. The criteria are based on the findings, using data from the 100-Car Naturalistic Driving Study, that crash risk increases when distraction durations exceed 2.00 seconds (Klauer et al., 2006). According to NHTSA’s (2012) guidelines, the following criteria should be met when testing new in-vehicle systems:

1. For at least 21 of the 24 test participants, the mean duration of all individual eye glances away from the forward road scene should be less than 2.00 seconds while performing the secondary task.

2. 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 should have durations of greater than 2.00 seconds while performing the secondary task.

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 should be less than, or equal to, 12.00 seconds while performing the secondary task one time.

The second criterion entails that for each driver performing a secondary task, no more than 15 percent of the off-road eye glances are allowed to be longer than 2.00 seconds, and only 3 in 24 drivers may fail to meet this criterion. The principle to consider the tail end of the distribution of off-road glance durations is in accordance with Horrey and Wickens (2007). They argued that the tail ends of the distributions of glance lengths are more sensitive to differences in crash risk than average glance durations. Wickens (2001) also argued that it is in the conditions in the tail ends of distributions that are likely to produce crashes rather than the mean, which represents typical conditions.

Since individual differences have been found to be an important part in explaining driving behavior it is likely that there are individual differences in visual scanning behavior while
driving. Ljung Aust, Dombrovskis, Kovaceva, Svanberg and Ivarsson (in press) have found that individual drivers display very varied visual attention shifting strategies when multi-tasking. Their work is based on an assumed interaction cycle when multi-tasking, see Figure 1. In the interaction cycle the driver checks safety margins in the traffic environment and then looks at a display, for example on an in-vehicle system on which a task is being performed. By doing this the driver updates his/her current understanding of both the driving task and the secondary task. The driver can then base the next actions on this current understanding of the situation, in accordance with COCOM (Hollnagel, 1999).

Figure 1
Typical interaction cycle when multi-tasking. Note that not each cycle will require an action; some cycles will be for information gathering only. (Ljung Aust et al., in press)

Ljung Aust et al. (in press) studied drivers’ off-road glance durations while the drivers performed five typical infotainment tasks during on road car driving. By investigating average off-road glance durations and the 85th percentile off-road glance duration, they found that the off-road glance durations were quite robust within each driver between tasks, but varied widely between drivers. This implies that there are individual differences between drivers regarding the amount of time they are willing to spend in each cycle phase in the interaction cycle.

They also found that it is not rare for drivers to have off-road glances longer than 2.00 seconds in more than 15 percent of the eye glances away from the forward roadway. This was the case even in simple tasks such as tuning the radio, which NHTSA (2012) has proposed as a reference task when testing in-vehicle systems because it is a commonly performed task that is socially acceptable to perform while driving. If the results of Ljung Aust et al. (in press) would be applied to the driving population in general, one in six drivers would have naturally long off-road glance durations. This implies that there are individuals who tend to have off-road glances longer than 2.00 seconds regardless of the design of the non-driving related in-vehicle systems.

Ljung Aust et al. (in press) suggested a reformulation of the second criterion in NHTSA’s guidelines, to avoid designs of non-driving related in-vehicle systems to be rejected because of natural variability among drivers. They suggested that compliance on some criteria in the guidelines might be better to measure on a group level rather than on an individual level, since
they interpreted the guidelines to have been derived using group level data. They noted, however, that all subjects in their study were Volvo Car employees and 8 of 35 subjects were directly involved in the design of visual-manual secondary task interfaces, which might introduce a bias. They used actual car driving in their study, while the guidelines propose that the criterion is tested in a driving simulator (NHTSA, 2012). The criterion should therefore be further investigated with subjects who better represent a general driving population and in a driving simulator.

The NHTSA guidelines also propose that the drivers only practice the tasks on the in-vehicle systems, which are to be tested, while the simulator is in parking mode. The combined driving and performing of secondary tasks in the test, is then a different task than performing the secondary task alone. If the drivers have not practiced the tasks while driving, it is likely that there is a learning effect when the tasks are performed several times while driving. Over the course of five sessions in a driving simulator, older drivers have been found to improve both driving performance while performing a distracting phone task and performance in the phone task (Shinar, Tractinsky, Compton, 2005). Younger drivers have also been found to improve driving performance in a simulator, while performing a task on an mp3-player, over the course of six sessions (Chisholm, Caird, Lockhart, 2008). It should therefore be investigated whether there is a learning effect when in-vehicle system tasks are performed several times in a car simulator, when drivers only have practiced the tasks while the simulator is in parking mode.
2. Purpose of the study
The purpose of the study was to investigate the 85\textsuperscript{th} percentile off-road glance durations when performing in-vehicle system secondary tasks while driving in a car simulator. The purpose was fourfold. The first part was to investigate the variation between individuals. The second part was to investigate whether there are differences between tasks. The third part was to investigate whether there is a learning effect when the tasks are performed at several occasions. The fourth part was to investigate whether there are any relations, which could explain potential individual differences in glance durations, between performance in the computerized trail-making test and the 85\textsuperscript{th} percentile off-road glance durations. That is, if factors such as ability of visual search, motor speed and mental flexibility are related with off-road glance durations.

The research questions were:

1a. How many individuals in the sample have an 85\textsuperscript{th} percentile off-road glance duration longer than 2.00 seconds when performing tasks on in-vehicle systems while driving in a car simulator?

1b. How does the distribution of individual drivers’ 85\textsuperscript{th} percentile off-road glance durations, when performing tasks on in-vehicle systems while driving in a car simulator, look?

2a. Regarding the 85\textsuperscript{th} percentile off-road glance durations when performing secondary tasks on in-vehicle systems while driving in a car simulator, are there any differences between three tasks and between the first, second and third occasion at which the tasks are performed?

2b. Are there any interaction effects between the tasks and at which occasion the tasks are performed?

3. Are there any correlations between performance in the computerized trail-making test created by Summala et al. (2008) and the 85\textsuperscript{th} percentile off-road glance durations when performing secondary tasks on in-vehicle systems while driving in a car simulator?
3. Method

3.1 Experimental design
A simulator study was performed at the Swedish National Road and Transport Research Institute, VTI. The study had a within-group design. All subjects performed three different in-vehicle system tasks while driving in a car simulator. They performed a radio task, a telephone task and a sound settings task. Each task was performed at three occasions by each of the subjects and all subjects performed the tasks in the same order, see Table 1.

Table 1
The order, in which the tasks were performed

<table>
<thead>
<tr>
<th>Order</th>
<th>Task</th>
<th>Occasion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>radio</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>telephone</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>sound settings</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>radio</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>telephone</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>sound settings</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>radio</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>telephone</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>sound settings</td>
<td>3</td>
</tr>
</tbody>
</table>

3.2 Subjects
30 subjects performed the test, but the results from 14 of them were not included in the study because of a significant amount of eye tracker data loss from their tests. Hence 16 subjects were included in the current study. The inclusion criteria, for participating in the study, were to be between 25 and 55 years old, to have had a driver’s license for at least five years, and to drive regularly, i.e. at least once a week. The subjects were recruited either through a list with people who had reported to VTI that they were willing to participate in studies, through acquaintances to the researcher and employees at VTI, or by asking employees at VTI themselves if they were willing to participate in the study. The subjects in the study did not work with design of visual-manual secondary task interfaces.

The 16 included subjects were 25 to 54 years old ($M = 36$, $SD = 8.6$). Nine of the subjects were men and seven were women. Three of the subjects drove less than 10 000 km per year, ten of the subjects drove 10 000 – 20 000 km per year, two of the subjects drove 20 000 – 30 000 km per year, and one of the subjects drove more than 30 000 km per year.

3.3 Apparatus
The simulator used in the study had a Saab 9-3 car compartment mounted on a motion platform. The platform allowed four degrees of freedom and offered both linear and tilt motion. The simulator also had a vibration table that was used to reproduce road unevenness and moved the compartment relative to the projected image, see Figure 2. The simulator’s visual system consisted of six high resolution projectors, which gave the driver a horizontal field of view of 115 degrees. Three LCD displays simulated rear-view mirrors.
A tablet computer, with a 9.7-inch touch screen, was positioned in front of the center stack in the car, see Figure 3. The computer was equipped with a prototype GUI of an infotainment system with several features, of which three was used in the study, i.e. phone, radio and sound settings.

A Smart Eye Pro 5.9 eye tracker system (Smart Eye, n.d.) was used to record the subjects’ eye gazes. Four IR cameras and two IR flashes were positioned in front of the driver. Three cameras were positioned on top of the dashboard and one was positioned just to the right of the tablet computer. One flash was positioned to the left of the dashboard and the other was positioned to the right of the dashboard. The eye tracker used the reflections of the flashes on the cornea to estimate the drivers’ gaze direction. A 3D model of the car compartment was used to analyze where the gaze was directed. The model included 2D planes that represented zones such as windshield, dashboard, center stack and tablet computer, see Figure 4. Data logging from the eye tracker was done at a sampling rate of 40 Hz.
After the simulator drive, the subjects performed an extended computerized version (Summala et al., 2008) of the trail-making test (Bowie & Harvey, 2006). The test was performed on a 23-inch touch screen.

3.4 Driving task and secondary tasks
The primary task was to drive on a motorway, with a signed speed limit of 90 km/h for approximately 30 minutes. The subjects were instructed to drive as they normally would do under similar circumstances, but to always try to maintain a speed of 90 km/h. The subjects were overtaken by other cars in average once every minute and they caught up with slower cars, which they had to overtake to be able to keep a speed at 90 km/h, in average once every two minutes.

During the drive the subjects performed three secondary tasks, which were executed on an in-vehicle system on the tablet computer, see Figure 3 in section 3.3. All secondary tasks were performed at three occasions, as shown in Table 1 in section 3.1, and were started on the in-vehicle system’s start page, see Figure 5.
Figure 5
The start page of the in-vehicle system. The red boxes highlight where the subjects tapped to start the tasks; from the top to the bottom as follows: the sound settings task, the radio task and the telephone task.

The first task was to tune in a specified radio frequency. By tapping items on the display, the subjects first opened the radio feature and then chose manual tuning. They tuned the frequency by dragging a marking on a frequency band to the specified frequency or by tapping arrows on each side of the frequency band, where each arrow tap corresponded to a 0.1 MHz change, see Figure 6. The frequencies that were to be tuned in alternated between 96.3 and 99.4 MHz.

Figure 6
The radio interface on the tablet computer

The second task was to dial their own 13 digit telephone number. By tapping items on the display, the subjects first opened the phone feature and then the keypad, on which they dialed their telephone number and tapped a call item, see Figure 7. They then hung up by tapping another item. Every time they opened the keypad, except for the first time, their number was already entered so they first had to delete it by holding a delete item for about a second or tap the delete item once for each figure. They were then able to enter the telephone number again. The subjects were informed that the infotainment system was a prototype and no signal would be sent to the receiving number.
The third task was to change the sound settings by changing the bass and treble levels. By tapping items on the display, the subjects opened the settings feature, the sound option and the equalizer. There were two bands, one for the bass level and one for the treble level, see Figure 8. The subjects changed the levels by dragging markings on the bands to the left, for lower levels, and to the right, for higher levels. The task alternated between changing the bass and treble levels to maximum and to minimum.

Figure 7
The telephone interface on the tablet computer

Figure 8
The sound settings interface on the tablet computer
The secondary tasks were announced by a pre-recorded voice message that said which task the subjects should perform. The voice message also told the details of the tasks, i.e. which radio frequency to tune in and whether the bass and treble levels were to be changed to maximum or minimum. Before the simulator drive had started, the subjects were instructed to tap a home item, after each secondary task, to go back to the in-vehicle system’s start page, so that they started each secondary task from the same page.

The current experiment also included an evaluation of a forward collision warning, FCW, system. A critical event, that triggered the FCW system, happened after approximately 20 minutes of driving and then again after 30 minutes of driving. It was important that the subjects did not look at the forward road when the critical event happened and an attention-shifting task was used to make the driver look away from the road. The task was to verbally repeat five digits that were shown on a display. This display was another display than the one on which the in-vehicle system tasks were performed. The attention-shifting task display was positioned to the right of the in-vehicle system display and of the center stack, as shown in Figure 3 in section 3.3. The subjects performed the attention-shifting task six times before the first critical event.

Data used in the present study was collected before the first critical event happened. The subjects performed the tasks in the following order: attention-shifting task, radio task, telephone task, sound settings task, attention-shifting task. This task sequence was repeated three times before the first critical event. The subjects performed an in-vehicle system task or attention-shifting task, in average once every minute, except for the three first minutes when no tasks were performed. The attention-shifting task was not performed during the in-vehicle system tasks. The onset of a task (both in-vehicle system tasks and attention-shifting tasks) was randomized in an interval of 50 seconds.

3.5 Trail-making test
To measure the subjects’ ability of visual search, motor speed and mental flexibility, the subjects performed a computerized trail-making test developed by Summala, et al. (2008). The test was performed on a 23-inch touch screen. In the test, the subjects tapped on targets, in the shape of circles with a width of 2.74 cm. The test consisted of two parts: part A, with targets marked with numbers, and part B, with targets marked with both numbers and letters. Figure 9 shows the screens for part A and part B, respectively. In part A, the targets were tapped in a numerical order, such as 1-2-3-4 etc., and in part B the targets were tapped in a combined numerical and alphabetical order, such as 1-A-2-B-3-C etc.
Figure 9
The screens for the trail-making test part A (left) and part B (right). Only the first target (1) was marked with a blue circle.

The circles were fixed on the screen but the content, i.e. numbers or letters, were either the same (fixed) or randomly changed after each tap. Hence, the test consisted of four sub-tests, which all were repeated once according to the order in Table 2. Each sub-test consisted of 20 targets. Before the test started the subjects trained on each type of sub-test with 12 targets.

Table 2
Order of the trail-making sub-tests

<table>
<thead>
<tr>
<th>Order</th>
<th>Round</th>
<th>Targets</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Part A fixed</td>
<td>numerical</td>
<td>fixed</td>
</tr>
<tr>
<td>2</td>
<td>Part A random</td>
<td>numerical</td>
<td>random</td>
</tr>
<tr>
<td>3</td>
<td>Part B fixed</td>
<td>numerical and alphabetical</td>
<td>fixed</td>
</tr>
<tr>
<td>4</td>
<td>Part B random</td>
<td>numerical and alphabetical</td>
<td>random</td>
</tr>
<tr>
<td>5</td>
<td>Part B random</td>
<td>numerical and alphabetical</td>
<td>random</td>
</tr>
<tr>
<td>6</td>
<td>Part B fixed</td>
<td>numerical and alphabetical</td>
<td>fixed</td>
</tr>
<tr>
<td>7</td>
<td>Part A random</td>
<td>numerical</td>
<td>random</td>
</tr>
<tr>
<td>8</td>
<td>Part A fixed</td>
<td>numerical</td>
<td>fixed</td>
</tr>
</tbody>
</table>

3.6 Procedure
The subjects were given written and spoken information about the study, about the test procedure, that their results were anonymous, and that they could terminate the test anytime they wanted without being asked for a reason. After an opportunity to ask questions they signed a consent form.

The subjects were shown to the simulator where they were given written instructions for the in-vehicle system tasks. They performed each task until they reported that they were comfortable performing them, while the simulator was in parking mode. Then a 10-minute training drive was run, so that the subjects grew accustomed to driving in the simulator. Next the actual driving began, during which the in-vehicle system tasks were performed.

After the simulator drive the subjects read written instructions for the trail-making test and performed four practice rounds with twelve circles, so that they practiced on each type of sub-test. They then performed the real test in which they performed the four sub-tests, two times
each, with 20 circles. Finally they answered a questionnaire with questions regarding what they thought about the difficulty of the in-vehicle system tasks and if they had had any problems performing the tasks.

### 3.7 Measures

Durations of the off-road glances when performing the in-vehicle system tasks while driving in the simulator were measured with an eye tracker. Off-road glances were defined as starting as soon as the driver’s gaze left the road and ended when the gaze returned to the road. The off-road glance durations were then used to calculate the 85th percentile off-road glance duration for each subject over all occasions at which each task was performed and for each subject at each occasion at which each task was performed. The 85th percentile off-road glance duration was calculated by rounding up, so that the measure represented the shortest duration in the 15 percent of the glances with the longest durations.

In the trail-making test the mean time between correct responses, i.e. the time between tapping two successive, correct targets, were calculated. The mean time between correct responses was calculated for each of the four sub-tests, i.e. part A fixed, part A random, part B fixed, and part B random. To isolate the ability of mental flexibility and control for the visual search component, the mean time between correct responses in part A fixed was subtracted from the mean time between correct responses in part B fixed. By subtracting mean time between correct responses in part A fixed from mean time between correct responses in part A random, the gain from having fixed targets were calculated, i.e. how much the subjects benefited from stable targets in relation to random targets. For the dependent variables in the study, see Table 3.

<table>
<thead>
<tr>
<th>Measure</th>
<th>For</th>
</tr>
</thead>
<tbody>
<tr>
<td>85th percentile off-road glance duration</td>
<td>each subject over all occasions at which each task was performed</td>
</tr>
<tr>
<td></td>
<td>each subject at each occasion at which each task was performed</td>
</tr>
<tr>
<td>mean time between correct responses</td>
<td>trail-making test, part A fixed</td>
</tr>
<tr>
<td></td>
<td>trail-making test, part A random</td>
</tr>
<tr>
<td></td>
<td>trail-making test, part B fixed</td>
</tr>
<tr>
<td></td>
<td>trail-making test, part B random</td>
</tr>
<tr>
<td></td>
<td>part B fixed – part A fixed</td>
</tr>
<tr>
<td></td>
<td>part A random – part A fixed</td>
</tr>
</tbody>
</table>

### 3.8 Data filtering

All analyses were performed on the data collected only during the in-vehicle system tasks. In-vehicle system tasks were defined as starting when the voice message, that announced the tasks, started. They were defined as ending when a new task (an in-vehicle system task or attention-shifting task) started or 60 seconds had passed since the start of the task. The length of the in-vehicle system tasks varied between 23 and 60 seconds.
The data from the eye tracking was not complete, and there was a substantial amount of data loss. To reduce errors due to incomplete data, all participants with a data loss of 20 percent or more were excluded from the study. Out of the 30 participants, who performed the test, 14 participants were excluded due to this criterion. The 16 remaining participants who were included in the study had a data loss of 3.01 to 19.07 percent ($M = 12.75$, $SD = 5.31$).

The eye tracker data consisted of rows containing information of what plane in the 3D model of the car compartment, see Figure 4 in section 3.3, that the gaze was directed at, i.e. the gaze target. Every row, i.e. data point, represented 0.025 seconds. A filter was used on this data to reduce noise and compensate for missing data. The filter calculated the most frequent gaze target in a window of eleven successive lines, so that a filtered value was presented at each row, i.e. every 0.025 seconds.

The in-vehicle system tasks involved reading and visual search. According to Rayner (1978), average fixation durations when reading were 0.20 to 0.25 seconds and mean fixation durations ranged from approximately 0.28 seconds up to 0.50 seconds in visual search experiments. Glances shorter than 0.20 seconds were therefore regarded physically implausible and were excluded in the analysis. Where the filter resulted in eight or more successive tablet computer gazes, i.e. tablet computer gazes equal to or longer than 0.20 seconds, they were marked as tablet computer glances. Tablet computer glances that were 3.00 seconds or longer were removed because of the risk of introducing measurement errors. Even though glances longer than 3.00 seconds are possible, it was decided that it was more important to avoid inflating the off-road glance durations because of measurement errors than to include these long glances. A 3.00 seconds limit has also been used in previous similar studies (Ljung Aust et al., in press).

The start and end of the glances were defined as five lines above the first and last tablet computer gaze, i.e. half the length of the filter window. If there were empty lines in the filtered data, i.e. no gazes that could be tracked, right before, right after or between tablet computer gazes, the empty lines were included in the glance duration.

Before the filter was applied, the data was processed according to the following rules. Behind the tablet computer plane in the 3D model of the car compartment was a larger plane representing the center stack, and behind that plane was a plane representing the whole dashboard, see Figure 4 in section 3.3. When the subjects looked from the windshield to the tablet computer their gaze passed over the dashboard and center stack planes. To include these gazes in the glances away from the forward road during in-vehicle system tasks, gazes logged as dashboard and center stack were replaced by tablet computer gazes.

Tablet computer gazes were regarded as more credible than other gaze targets during the in-vehicle system tasks, and the filter therefore weighted rows with tablet computer gazes by 2.0. The rows, where no gaze had been tracked, were weighted by 0.3 because it was regarded credible that these gazes were directed at the same object as the gazes that had been tracked right before or after. For an example of the data filtering, see Appendix A.
3.9 Statistical analysis

The 85\textsuperscript{th} percentile off-road glance durations for each driver and each task were calculated and durations longer than 2.00 seconds were marked. Also, a graph was plotted to show the variation in the 85\textsuperscript{th} percentile off-road glance duration between drivers and between tasks within each driver.

A factorial 3*3 ANOVA with repeated measures was used to analyze if there were any differences in 85\textsuperscript{th} percentile off-road glance durations between the three in-vehicle system tasks (radio task, telephone task and sound settings task) and between the first, second and third occasion at which the tasks were performed. The interaction effect between in-vehicle system task and at which occasion the tasks was performed was also calculated. Bonferroni post-hoc tests were used to investigate differences further.

Two-tailed Pearson correlations between the 85\textsuperscript{th} percentile off-road glance duration for each task and the mean response time in each of the trail-making sub-tests, B fixed minus A fixed and A random minus A fixed, were calculated. For all statistical analyses in the study a 5-percent level of significance was used.
4. Results

4.1 Variation in 85th percentile off-road glance durations

The mean number of glances away from the road, for each occasion at which each task was performed, for the total of each task and for the total of each occasion, are presented in Table 4.

Table 4
Number of off-road glances

<table>
<thead>
<tr>
<th>Occasion</th>
<th>Task</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>radio</td>
<td>14.00</td>
<td>3.46</td>
</tr>
<tr>
<td>1</td>
<td>telephone</td>
<td>14.31</td>
<td>4.39</td>
</tr>
<tr>
<td></td>
<td>sound settings</td>
<td>14.81</td>
<td>3.83</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>43.13</td>
<td>8.16</td>
</tr>
<tr>
<td>2</td>
<td>radio</td>
<td>12.06</td>
<td>4.15</td>
</tr>
<tr>
<td></td>
<td>telephone</td>
<td>13.88</td>
<td>3.96</td>
</tr>
<tr>
<td></td>
<td>sound settings</td>
<td>11.56</td>
<td>3.63</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>37.50</td>
<td>8.01</td>
</tr>
<tr>
<td>3</td>
<td>radio</td>
<td>11.63</td>
<td>3.34</td>
</tr>
<tr>
<td></td>
<td>telephone</td>
<td>11.19</td>
<td>3.25</td>
</tr>
<tr>
<td></td>
<td>sound settings</td>
<td>10.13</td>
<td>3.14</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>32.94</td>
<td>8.25</td>
</tr>
<tr>
<td>1, 2 and 3</td>
<td>radio</td>
<td>37.69</td>
<td>7.62</td>
</tr>
<tr>
<td></td>
<td>telephone</td>
<td>39.38</td>
<td>8.63</td>
</tr>
<tr>
<td></td>
<td>sound settings</td>
<td>36.50</td>
<td>8.77</td>
</tr>
</tbody>
</table>

The 85th percentile off-road glance durations for each subject over all occasions at which each task was performed are presented in Table 5. One subject had an 85th percentile off-road glance duration longer than 2.00 seconds in all three tasks. Two subjects had an 85th percentile off-road glance duration longer than 2.00 seconds in the telephone and sound settings tasks. One subject had an 85th percentile off-road glance duration longer than 2.00 seconds in the radio and telephone tasks. Three subjects had an 85th percentile off-road glance duration longer than 2.00 seconds in one of the tasks, two of them were in the telephone task and one of them was in the sound settings task. The means of the 85th percentile off-road glance durations for the three tasks, for each subject, are also presented in Table 5.
Table 5
The 85th percentile off-road glance durations, in seconds, for each subject over all occasions at which each task was performed. The mean of the 85th percentiles for all three tasks for each subject is also presented. The subjects are ordered by this mean.

<table>
<thead>
<tr>
<th>subject</th>
<th>radio</th>
<th>telephone</th>
<th>sound settings</th>
<th>mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.05</td>
<td>1.48</td>
<td>1.25</td>
<td>1.26</td>
</tr>
<tr>
<td>2</td>
<td>1.20</td>
<td>1.40</td>
<td>1.28</td>
<td>1.29</td>
</tr>
<tr>
<td>3</td>
<td>1.40</td>
<td>1.48</td>
<td>1.43</td>
<td>1.43</td>
</tr>
<tr>
<td>4</td>
<td>1.40</td>
<td>1.58</td>
<td>1.35</td>
<td>1.44</td>
</tr>
<tr>
<td>5</td>
<td>1.48</td>
<td>1.60</td>
<td>1.43</td>
<td>1.50</td>
</tr>
<tr>
<td>6</td>
<td>1.40</td>
<td>1.65</td>
<td>1.53</td>
<td>1.53</td>
</tr>
<tr>
<td>7</td>
<td>1.40</td>
<td>1.70</td>
<td>1.70</td>
<td>1.60</td>
</tr>
<tr>
<td>8</td>
<td>1.58</td>
<td>1.80</td>
<td>1.80</td>
<td>1.73</td>
</tr>
<tr>
<td>9</td>
<td>1.65</td>
<td>1.85</td>
<td>1.90</td>
<td>1.80</td>
</tr>
<tr>
<td>10</td>
<td>1.60</td>
<td>2.10</td>
<td>1.70</td>
<td>1.80</td>
</tr>
<tr>
<td>11</td>
<td>1.63</td>
<td>2.08</td>
<td>1.83</td>
<td>1.84</td>
</tr>
<tr>
<td>12</td>
<td>1.80</td>
<td>1.90</td>
<td>2.13</td>
<td>1.94</td>
</tr>
<tr>
<td>13</td>
<td>2.25</td>
<td>2.05</td>
<td>1.70</td>
<td>2.00</td>
</tr>
<tr>
<td>14</td>
<td>1.88</td>
<td>2.18</td>
<td>2.05</td>
<td>2.03</td>
</tr>
<tr>
<td>15</td>
<td>1.95</td>
<td>2.08</td>
<td>2.13</td>
<td>2.05</td>
</tr>
<tr>
<td>16</td>
<td>2.55</td>
<td>2.40</td>
<td>2.55</td>
<td>2.50</td>
</tr>
</tbody>
</table>

Figure 10 presents the 85th percentile off-road glance duration for each subject over all occasions at which each task was performed. The graph shows variation between subjects and between tasks within each subject. The subjects are ordered by the mean of the 85th percentile off-road glance durations for the three tasks for each subject.

Figure 10
85th percentile off-road glance durations for the three in-vehicle system tasks for each subject, ordered by the mean 85th percentile off-road glance duration for all tasks

4.2 Differences between tasks and occasions
Mauchly’s test indicated that sphericity could be assumed for task, $\chi^2(2) = 3.55, p = 0.170,$ and the interaction of task and at which occasion (1, 2 or 3) the tasks were performed, $\chi^2(9) = 7.61, p = 0.577.$ The factorial ANOVA showed that there was a significant main effect of
task, $F(2, 30) = 15.34, p < 0.050$, but no significant interaction effect of task and at which occasion the tasks were performed, $F(4, 60) = 0.79, p = 0.537$. Mauchly's test indicated that the assumption of sphericity had been violated for at which occasion the tasks were performed, $\chi^2(2) = 6.01, p = 0.049$. Therefore the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.74$). The factorial ANOVA then showed that there was no significant main effect of at which occasion the tasks were performed, $F(1.48, 22.24) = 3.46, p = 0.061$.

A Bonferroni post-hoc test showed significant differences between the radio task and the other two tasks, $p < 0.050$. The radio task had lower 85\textsuperscript{th} percentile off-road glance durations than the telephone task and sound settings task, see Figure 11.

![Figure 11](image11.png)

**Figure 11**
Mean 85\textsuperscript{th} percentile off-road glance durations for the three tasks

Even though there was no significant main effect of at which occasion the tasks were performed, the $F$-value was close to being significant. Figure 12 shows that the mean 85\textsuperscript{th} percentile off-road glance durations for all three tasks decreased over the occasions, at which the subjects performed the tasks.

![Figure 12](image12.png)

**Figure 12**
Mean 85\textsuperscript{th} percentile off-road glance durations for the three occasions
There was no significant interaction effect between task and at which occasion the task was performed. Figure 13 shows the mean 85\textsuperscript{th} percentile off-road glance duration for each task at each occasion.

The 85\textsuperscript{th} percentile off-road glance durations have similar tendencies in the telephone task and sound settings task, for which the durations decrease over occasion. The radio task does not show the same pattern, but rather shows a tendency of being quite stable across the occasions. Therefore a factorial 2*3 ANOVA with repeated measures was used to analyze if there were any differences in 85\textsuperscript{th} percentile off-road glance duration between the telephone and sound settings tasks and between the first, second and third occasion at which these tasks were performed. There was also analyzed if there was an interaction effect between these tasks and at which occasion they were performed.

Mauchly’s test indicated that sphericity could be assumed for occasion $\chi^2 (2) = 2.50$, $p = 0.287$, and the interaction of task and at which occasion the tasks were performed, $\chi^2 (2) = 1.71$, $p = 0.426$. There was a significant main effect of task, $F(1, 15) = 6.32$, $p < 0.050$, and of occasion, $F(2, 30) = 3.39$, $p < 0.050$. There was no significant interaction effect, $F(2, 30) = 0.32$, $p = 0.729$.

A Bonferroni post-hoc test showed no significant differences between the three occasions, but a tendency of a difference between occasion 1 and occasion 3 was found, $p = 0.068$. An LSD (least significant difference) post-hoc test showed a significant difference between occasion 1 and occasion 3, $p < 0.050$.

4.3 Individual differences and 85\textsuperscript{th} percentile off-road glance durations

There were no significant correlations between the mean response times in the trail-making sub-tests, in part B fixed minus part A fixed or in part A random minus part A fixed and the 85\textsuperscript{th} percentile off-road glance durations for the three in-vehicle system tasks, see Table 6. Figure 14 shows the mean response time for each subject in each sub-test.
Table 6
2-tailed Pearson correlations between mean response times in the trail-making sub-tests, in part B fixed minus part A fixed, in part A random minus part A fixed and 85th percentile off-road glance durations in the in-vehicle system tasks, N = 16

<table>
<thead>
<tr>
<th></th>
<th>radio</th>
<th>telephone</th>
<th>sound settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>A fixed</td>
<td>Correlation</td>
<td>0.063</td>
<td>-0.097</td>
</tr>
<tr>
<td></td>
<td>Sig.</td>
<td>0.817</td>
<td>0.720</td>
</tr>
<tr>
<td>A random</td>
<td>Correlation</td>
<td>0.172</td>
<td>-0.096</td>
</tr>
<tr>
<td></td>
<td>Sig.</td>
<td>0.524</td>
<td>0.723</td>
</tr>
<tr>
<td>B fixed</td>
<td>Correlation</td>
<td>-0.102</td>
<td>-0.332</td>
</tr>
<tr>
<td></td>
<td>Sig.</td>
<td>0.706</td>
<td>0.210</td>
</tr>
<tr>
<td>B random</td>
<td>Correlation</td>
<td>-0.099</td>
<td>-0.346</td>
</tr>
<tr>
<td></td>
<td>Sig.</td>
<td>0.716</td>
<td>0.189</td>
</tr>
<tr>
<td>B fixed – A fixed</td>
<td>Correlation</td>
<td>-0.247</td>
<td>-0.467</td>
</tr>
<tr>
<td></td>
<td>Sig.</td>
<td>0.357</td>
<td>0.069</td>
</tr>
<tr>
<td>A random – A fixed</td>
<td>Correlation</td>
<td>0.065</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>Sig.</td>
<td>0.812</td>
<td>0.881</td>
</tr>
</tbody>
</table>

Figure 14
Mean response times for the four trail-making sub-tests for each subject, ordered by the mean response time for all sub-tests

It was not in the purpose of the study to investigate relations between age and the 85th percentile off-road glance durations. When investigating the data, significant positive correlations between age and the 85th percentile off-road glance duration for each task was however found, see Table 7. The older the drivers were, the longer 85th percentile off-road glance durations they had.

Table 7
2-tailed Pearson correlations between age and 85th percentile off-road glance durations in the in-vehicle system tasks, N = 16

<table>
<thead>
<tr>
<th>age</th>
<th>radio</th>
<th>telephone</th>
<th>sound settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td>0.664</td>
<td>0.635</td>
<td>0.676</td>
</tr>
<tr>
<td>Sig.</td>
<td>0.005</td>
<td>0.008</td>
<td>0.004</td>
</tr>
</tbody>
</table>
4.4 Glances at the forward road
The eye tracker was reliable at recording glances directed to the forward road because three of the eye tracker cameras were positioned right below the windshield. There was only one eye tracker camera positioned near the tablet computer and therefore the most data loss was when the subjects looked at the tablet computer. Since glances through the windshield were reliably recorded, it was investigated if there was a difference between the subjects, which were included in the analysis, and the subjects, which were excluded from the analysis, in how much they looked through the windshield at the forward road while performing the in-vehicle system tasks.

The subjects, which were included in the analysis, looked through the windshield in an average of 54.67 percent (SD = 8.88) of the data points. The subjects, which were excluded from the analysis, looked through the windshield in an average of 52.01 percent (SD = 7.26) of the data points. An independent samples t-test showed no significant difference between the included and excluded subjects, t(28) = 0.89, p = 0.382.

4.5 Questionnaire
Out of the included subjects three reported that they had a problem, with managing the in-vehicle system tasks, which was due to unfamiliarity with the tasks. Two of them reported that they had difficulties with navigating in the menu and one of them reported that s/he had problems because s/he was not accustomed to the system.

How difficult the subjects thought the in-vehicle system tasks were, was measured on a 7-point scale for all three tasks combined, on which 1 represented easy and 7 represented difficult. Subjective difficulty of the tasks was significantly correlated with the 85th percentile off-road glance durations for the radio task (Pearson correlation = 0.59, p < 0.050) and sound settings task (Pearson correlation = 0.59, p < 0.050). Subjective difficulty of the tasks was not correlated with the 85th percentile off-road glance durations in the telephone task (Pearson correlation = 0.41, p = 0.116).
5. Discussion

5.1 Variation in 85th percentile off-road glance durations
The distribution of 85th percentile off-road glance durations among individual drivers showed that there was a variation in glance durations between individuals. The 85th percentile off-road glance durations ranged from 1.05 seconds to about 2.55 seconds. The distribution also showed that there was only a small variation between the tasks within each subject, so that if a subject had a long 85th percentile off-road glance duration in one of the tasks, s/he was likely to have long 85th percentile off-road glance durations in the other tasks too.

This indicates that drivers use individual strategies when it comes to how long they are willing to look away from the road while performing tasks on in-vehicle systems, and that type of task does not substantially affect this strategy. There are drivers who feel comfortable looking away from the road for longer periods and drivers who only feel comfortable looking away from the road during shorter periods.

Out of the 16 subjects, 7 had an 85th percentile off-road glance duration longer than 2.00 seconds in at least one of the tasks. This shows that it is not unusual for drivers to look away from the road longer than 2.00 seconds while performing secondary tasks. Two, i.e. 12.50 percent, of the subjects had an 85th percentile off-road glance duration longer than 2.00 seconds in the radio task. Radio tuning is a commonly performed task that is socially acceptable to perform while driving and NHTSA (2012) considered the task to be acceptable to perform while driving. The results in this study do however imply that there exist drivers who feel comfortable to look away from the road longer than 2.00 seconds while tuning in a radio. If the results in this study are generalizable to a sample of 24 subjects, 3 of them, i.e. 12.50 percent, would have an 85th percentile off-road glance duration longer than 2.00 seconds. This means that the radio tuning task only barely would pass NHTSA’s proposed criterion for acceptance of in-vehicle systems, which proposes that for at least 21 of 24 test participants, no more than 15 percent of the total number of eye glances away from the forward road scene should have durations longer than 2.00 seconds while performing a secondary task. This criterion does therefore not comply with the premise that radio tuning is acceptable to perform while driving.

5.2 Differences between occasions
The 3*3 ANOVA showed no significant differences in 85th percentile off-road glance duration between the occasions at which the subjects performed the tasks. The F-value was not significant with Greenhouse-Geisser correction, but the p-value was quite low, i.e. 0.061. A graph with mean 85th percentile off-road glance durations of the tasks for each occasion showed that there was a tendency that the 85th percentile off-road glance durations decreased over the three occasions, so that glances became shorter for each occasion. Hence, there was a tendency of a learning effect. When performing tasks at more than three occasions it is probable that a significant difference would emerge. This tendency implies that the 85th percentile off-road glance durations decrease when drivers get more familiar with the task. As the tasks were performed at several occasions, the subjects had fewer off-road glances. This strengthens the tendency that there was a learning effect.
The 3*3 ANOVA did not show a significant interaction effect between task and occasion, regarding 85\textsuperscript{th} percentile off-road glance durations. However, a graph, shown in Figure 13 in section 4.2, which presents the mean 85\textsuperscript{th} percentile off-road glance durations for each task at each occasion, showed tendencies for the 85\textsuperscript{th} percentile off-road glance duration to decrease over the three occasions for the telephone task and sound settings task. The radio task did not show the same tendency. The 85\textsuperscript{th} percentile off-road glance durations in the radio task were quite stable across the three occasions. The radio task had significantly shorter 85\textsuperscript{th} percentile glance durations than the other two tasks. There might therefore be a floor effect that entails that 85\textsuperscript{th} percentile off-road glance durations do not get shorter than about 1.55 seconds, while performing tasks on in-vehicle systems while driving. It is possible that the other tasks would have reached this floor value if the drivers would have performed the tasks more times and thus had become more familiarized with the tasks. This possible floor effect should be further investigated.

Because the radio task showed a different tendency than the other two tasks, a 2*3 ANOVA with only the telephone and sound settings tasks over the three occasions were analyzed. This analysis showed that there was a significant main effect of occasion. Post-hoc tests showed a tendency that there were longer 85\textsuperscript{th} percentile off-road glance durations at the first occasion than at the third occasion, at which the two tasks were performed. This implies that glances in tasks, which demand longer glances, decrease when the driver gets more familiar with performing the tasks while driving.

These results suggests that it is likely that 85\textsuperscript{th} percentile off-road glance durations differ between tasks the first times they are performed, but that the difference decrease when they are performed several times. When drivers are totally familiar with performing the tasks it is possible that this difference ceases to exist. Further studies should investigate differences in 85\textsuperscript{th} percentile off-road glance durations, when performing secondary tasks while driving, when the tasks are performed at more than three occasions and when the subjects are totally familiar with the tasks.

5.3 Differences between tasks

Even though the distribution of 85\textsuperscript{th} percentile off-road glance durations for each subject and task showed that the variation in durations for each task within subjects was small, ANOVA showed that there were differences between the durations in the tasks on a group level. The radio task generated shorter 85\textsuperscript{th} percentile off-road glance durations than the telephone task and the sound settings task, according to the 3*3 ANOVA, and the sound settings task generated shorter 85\textsuperscript{th} percentile off-road glance durations than the telephone task, according to the 2*3 ANOVA, in which the radio task was not included. This implies that even if drivers use different strategies of how long they look away from the road, there are also differences between tasks on a group level, at least before the drivers are familiar with performing the tasks while driving. These differences could be the result of the design of the task.

The radio task required two taps to navigate to a frequency band, on which a marking was dragged to a specified frequency, and two taps to return to the start page. The marking could also be moved by tapping arrows on each side of the frequency band. The telephone task
required two taps to navigate to a keypad, then 13 more taps on the keypad, and finally one tap to end the call and to return to the start page. The sound settings task required four taps to navigate to a sound settings page, where markings on two bands were dragged to maximum or minimum, and one tap to return to the start page. The two tasks that required the most steps, i.e. the telephone and sound settings task, had longer off-road glance durations than the radio task, which required fewer steps. This may imply that tasks that require many steps results in longer glances away from the road.

In the telephone task, most of the items to be tapped were on a keypad. Most subjects should have been familiar to the design of the keypad from telephones, computers, calculators etcetera. The task should therefore not have required as much visual search and memory as the sound settings task. In the sound settings task the subjects had to navigate through menus to find the bass level and treble level bands, which should have demanded memory, to remember how to navigate through the menu, and visual search, to find the items to tap in the menu. Even though the sound settings task should require more visual search and memory, it resulted in shorter off-road glance durations than the telephone task according to the 2*3 ANOVA. This may imply that the number of taps is more critical than the need of visual search and memory to find the items to tap, when it comes to off-road glance durations.

The 85th percentile off-road glance durations in the radio and sound settings tasks were correlated with how difficult the subjects found the in-vehicle system tasks. This implies that difficult tasks generate long off-road glance durations. The 85th off-road glance durations in the telephone task was however not correlated with subjective difficulty. It should be investigated what in the design of in-vehicle tasks that causes long off-road glance durations and if a small number of steps, to perform tasks, decrease off-road glance durations. Knowledge about how the design of tasks affects the interaction with the driver would help to create in-vehicle systems that are acceptable to use while driving and that meet criteria regarding safety, such as the criteria in NHTSA’s guidelines.

5.4 Trail-making test and 85th percentile off-road glance durations
There was no relation between performance in the trail-making test and 85th percentile off-road glance durations. There were no correlations between off-road glance durations when performing the in-vehicle system tasks and performance in any of the sub-tests. Nor were there any correlations between the off-road glance durations and mental flexibility controlled for visual search, i.e. B fixed minus A fixed, and the gain from stable targets, i.e. A random minus A fixed. The trail-making test is supposed to measure visual search, motor speed and mental flexibility. The results indicate that these factors do not affect how long drivers look away from the road while performing in-vehicle system tasks, even though the tasks are both visual and motoric and should demand mental flexibility.

Put differently, individuals seem to have different visual strategies when it comes to how long they are willing to look away from the road while performing tasks on in-vehicle systems, and these differences cannot be explained by performance in the trail-making test, i.e. visual search, motor speed and mental flexibility. There was not a big variation in performance in the trail-making sub-tests between subjects. It therefore seems that drivers in the age span
used in this study, i.e. 25 to 55 years old, do not vary substantially in ability of visual search, motor speed and mental flexibility.

One possible explanation for the lack of correlation is that the trail-making test is a primary task, where the subjects can direct all their attention on a single task to achieve maximum performance. The tasks on the in-vehicle systems on the other hand were secondary tasks, where the subjects had to shift attention between the driving task and the in-vehicle system tasks. Hence, even though a driver is good at visual search and has high motor speed and mental flexibility, s/he still seems to interact based on an individual attention sharing strategy that has little or nothing to do with his/her maximum performance capacity.

5.5 Age and 85th percentile off-road glance durations

In further analysis of the data it was found that age was positively correlated with the 85th percentile off-road glance duration in each of the three in-vehicle system tasks. This may imply that older drivers are more prone to have long off-road glance durations than younger drivers. Age may therefore be an individual factor that affects the trade-off between driving and performing tasks on in-vehicle systems. It was not planned to investigate correlations between age and off-road glance duration. It must therefore be further investigated if there exists such a relation.

Since the performance on the trail-making test was not related with the off-road glance durations, how long drivers are willing to look away from the road while performing in-vehicle system tasks must be due to something else. Even if there is a relation between age and off-road glance durations it would still be unclear why old drivers are more willing to have long off-road glance durations. It should therefore also be further investigated what causes some individuals to be more willing to look away from the road in longer periods than others.

5.6 Limitations

In NHTSA’s guidelines (2012) it was suggested that tests on in-vehicle systems and glances away from the road while using the systems, are performed with 24 subjects. In the present study, while 30 subjects participated, 14 had to be excluded from the analysis due to significant eye tracker data loss. If data from these subjects would have been used in the study the results would have been unreliable because quite major assumptions, of where the subjects looked when no data was logged, would have been assumed. There was however no difference between how much the included and excluded subjects looked through the windshield at the forward road while performing the in-vehicle system tasks. It is therefore likely that the excluded subjects had similar glance behavior as the included subjects. In future studies, the positions of the eye tracker cameras and flashes should be carefully considered to decrease data loss when the subjects look down at the in-vehicle system.

The subjects performed the in-vehicle system tasks in the same order for each repetition of the task set, which could result in an order effect. When the tasks were performed was, however, randomized so that they were not performed at the same time or at the same place, which should decrease the risk of an order effect. Since there was a tendency that 85th percentile off-
road glance durations decreased as the tasks were performed several times, there could be an order effect in the same direction between the tasks. Then the task that was performed first, i.e. the radio task, would have too long glance durations relative to the other two tasks. If this is the case, there should still be a difference between the radio task and the other two tasks when a possible order effect is accounted for, but it could be an even bigger difference.

That the in-vehicle systems tasks were started randomly in an interval of 50 seconds resulted in that the subjects had different amounts of time to finish the tasks until a new task was started. This might have affected the results if the glance durations in the beginning of the tasks are different from the glance durations in the end of the tasks. The amount of time they had to finish a task was however random, so this should not have affected the results.

Even though the subjects had practiced on the in-vehicle system tasks before driving in the simulator, three of the subjects reported that they had problems performing the tasks because they were not accustomed to the system. In future studies, it should be ensured that the subjects know how to perform the tasks before driving in the simulator. If possible in future studies, the steps that the subjects take to perform the tasks should be logged. Then it would be clear when the subjects are finished with the tasks and if they made any errors while performing them.

That the subjects had problems performing the tasks may however be due to the fact that they performed the tasks while driving. The three subjects reported that they were comfortable performing the tasks while focusing solely on them but when the tasks were combined with driving they had problems performing them. This implies that the subjects should practice on the tasks while driving before the test.

To make the data useable, it had to be filtered to reduce noise and to compensate for loss of data. A filter manipulates the data, which in turn affects the results. The filter must therefore be adjusted to make the data reflect reality as much as possible. The filter was based on assumptions about where the subjects were most likely to look while performing the in-vehicle system tasks. When no data was logged around a glance on the tablet computer, on which the tasks were performed, it was assumed that the subjects looked at the tablet computer. This was assumed because three eye tracker cameras were positioned in front of the driver, right below the windshield, and they were reliable at recording glances at the forward road. Gazes through the windshield were therefore not likely to be unlogged. It was very probable that gazes that were not logged were directed at the tablet computer because the subjects had to look at the tablet computer to perform the tasks.

The assumption that the subjects looked at the tablet computer when no data had been logged around glances at the tablet computer entails a risk of false positives. False positives would result in too long 85th percentile off-road glance durations. To decrease the risk of false positives and too long durations, the data was compared with video recordings from the test and results from previous studies investigating durations of off-road glances (Ljung Aust et al., in press). The comparisons showed that the filter resulted in credible data.
6. Conclusion

The duration of glances away from the road is an important factor when it comes to safety when performing tasks on in-vehicle systems while driving. When the driver looks away from the forward road s/he is not able to update his/her current understanding about the traffic situation. Glances away from the road that are longer than 2.00 seconds have been shown to be dangerous (Klauer et al., 2006; Ryu, Sihn & Yu, 2013). NHTSA (2012) has proposed guidelines when testing in-vehicle systems, to control that they are acceptable to use while driving.

One of the criteria in the guidelines is that for at least 21 of 24 test participants, no more than 15 percent of the total number of eye glances away from the forward road scene should have durations longer than 2.00 seconds when performing a secondary task while driving in a simulator. This criterion concerns the 85th percentile off-road glance duration for individual drivers when they perform secondary tasks. To look at the tail ends of distributions like this when regarding safety, is useful since accidents are more likely to happen in the extreme conditions in the tail ends of distributions than in normal conditions (Horrey & Wickens, 2007; Wickens, 2001).

The present study has however shown that 85th percentile off-road glance durations, when performing three different tasks while driving in a simulator, vary between individuals rather than between different tasks within individuals. It seems that drivers use different strategies, regarding off-road glance durations when performing the tasks. These strategies seem to be more affected by individual differences than tasks. It was not unusual with 85th percentile off-road glance durations longer than 2.00 seconds and two subjects had 85th percentile off-road glance durations longer than 2.00 seconds in a radio tuning task. These findings problematize NHTSA’s (2012) criterion, since radio tuning is considered to be a task that is acceptable to perform while driving and should therefore pass the criterion without any problems.

There were however differences between the tasks on a group level. The radio task had shorter 85th percentile off-road glance durations than the telephone task and sound settings task, and the sound settings task had shorter 85th percentile off-road glance durations than the telephone task. This shows that even though there were small variations between tasks within individuals, the differences were consistent which led to differences on a group level. A criterion for testing 85th percentile off-road glance durations when using in-vehicle systems might therefore be more appropriate to be based on group level data.

Tendencies of a learning effect, which resulted in a decrease of the 85th percentile off-road glance durations when tasks were performed several times, were found. The testing of in-vehicle systems should resemble how the systems will be used in actual driving. With the assumption that drivers regularly use the in-vehicle systems that they have in their cars, subjects should be familiar with the systems when testing them. Subjects should therefore practice on the tasks on the in-vehicle systems while driving in the simulator until they are familiarized with the task and in-vehicle system before the testing starts. A tendency of a floor effect, which implies that tasks reach a floor value when the driver is familiarized with the
tasks, was however found. Even though differences between tasks were found in this study, these differences could cease to exist if the subjects are familiarized with the tasks.

Drivers’ strategies for how long they are willing to look away from the road while performing secondary tasks on in-vehicle systems are not related with performance in the computerized trail-making test created by Summala et al. (2008). How long drivers are willing to look away from the road is therefore not affected by the drivers’ capacity in visual search, motor speed and mental flexibility, and should rather be affected by priorities in trade-offs between the driving task and performing the in-vehicle system task.
References


## Appendix A: Data filtering

<table>
<thead>
<tr>
<th>timer (s)</th>
<th>original data</th>
<th>filtered data</th>
</tr>
</thead>
<tbody>
<tr>
<td>320.000</td>
<td>windshield</td>
<td>windshield</td>
</tr>
<tr>
<td>320.025</td>
<td>windshield</td>
<td>windshield</td>
</tr>
<tr>
<td>320.050</td>
<td>windshield</td>
<td>windshield</td>
</tr>
<tr>
<td>320.075</td>
<td>windshield</td>
<td>windshield</td>
</tr>
<tr>
<td>320.100</td>
<td>windshield</td>
<td>windshield</td>
</tr>
<tr>
<td>320.125</td>
<td>windshield</td>
<td>windshield</td>
</tr>
<tr>
<td>320.150</td>
<td>windshield</td>
<td>windshield</td>
</tr>
<tr>
<td>320.200</td>
<td>windshield</td>
<td>windshield</td>
</tr>
<tr>
<td>320.225</td>
<td>windshield</td>
<td>windshield</td>
</tr>
<tr>
<td>320.250</td>
<td>tablet computer</td>
<td>windshield</td>
</tr>
<tr>
<td>320.275</td>
<td>windshield</td>
<td>windshield</td>
</tr>
<tr>
<td>320.300</td>
<td>tablet computer</td>
<td>windshield*</td>
</tr>
<tr>
<td>320.325</td>
<td>tablet computer</td>
<td>tablet computer**</td>
</tr>
<tr>
<td>320.350</td>
<td>tablet computer</td>
<td>tablet computer***</td>
</tr>
<tr>
<td>320.375</td>
<td>tablet computer</td>
<td>tablet computer</td>
</tr>
<tr>
<td>320.400</td>
<td>tablet computer</td>
<td>tablet computer</td>
</tr>
</tbody>
</table>

The filter calculates the most frequent gaze target in a window of eleven successive lines, by weighting tablet computer glances by 2.0 and empty data points by 0.3.

The tablet computer glance starts five rows above the first filtered tablet computer gaze (half the length of the filter window), i.e. at 320.200 s.

* **windshield**: \(1 \times 6 = 6\)
  * <empty>: \(0.3 \times 3 = 0.9\)
  * tablet computer: \(2 \times 2 = 4\)

** **windshield**: \(1 \times 5 = 5\)
  * <empty>: \(0.3 \times 3 = 0.9\)
  * tablet computer: \(2 \times 3 = 6\)

*** **windshield**: \(1 \times 4 = 4\)
  * <empty>: \(0.3 \times 3 = 0.9\)
  * tablet computer: \(2 \times 4 = 8\)
Upphovsrätt

Detta dokument hålls tillgängligt på Internet – eller dess framtida ersättare – från publiceringsdatum under förutsättning att inga extraordinära omständigheter uppstår.

Tillgång till dokumentet innebär tillstånd för var och en att läsa, ladda ner, skriva ut enstaka kopior för enskilt bruk och att använda det oförändrat för icke-komersiell forskning och för undervisning. Överföring av upphovsrätten vid en senare tidpunkt kan inte upphäva detta tillstånd. All annan användning av dokumentet kräver upphovsmannens medgivande. För att garantera äktheten, säkerheten och tillgängligheten finns lösningar av teknisk och administrativ art.

Upphovsmannens ideella rätt innefattar rätt att bli nämnt som upphovsman i den omfattning som god sed kräver vid användning av dokumentet på ovan beskrivna sätt samt skydd mot att dokumentet ändras eller presenteras i sådan form eller i sådant sammanhang som är kränkande för upphovsmannens litterära eller konstnärliga anseende eller egenart.

För ytterligare information om Linköping University Electronic Press se förlagets hemsida http://www.ep.liu.se/

Copyright

The publishers will keep this document online on the Internet – or its possible replacement – from the date of publication barring exceptional circumstances.

The online availability of the document implies permanent permission for anyone to read, to download, or to print out single copies for his/hers own use and to use it unchanged for non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional upon the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility.

According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement.
For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its www home page: http://www.ep.liu.se/

© Linnea Wahlberg