

**This is an electronic reprint of the original article.
This reprint *may differ* from the original in pagination and typographic detail.**

Author(s): Kujala, Tuomo; Saariluoma, Pertti

Title: Effects of menu structure and touch screen scrolling style on the variability of glance durations during in-vehicle visual search tasks

Year: 2011

Version:

Please cite the original version:

Kujala, T., & Saariluoma, P. (2011). Effects of menu structure and touch screen scrolling style on the variability of glance durations during in-vehicle visual search tasks. *Ergonomics*, 54(8), 716-732. <https://doi.org/10.1080/00140139.2011.592601>

All material supplied via JYX is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Effects of menu structure and touch screen scrolling style on the variability of glance durations during in-vehicle visual search tasks

Tuomo Kujala*

Agora Human Technology Center, University of Jyväskylä, Finland

P.O. Box 35, FIN-40014 Jyväskylä, Finland; e-mail: tuomo.kujala@jyu.fi,
tel. +358 14 260 4660, fax +358 14 260 4400

Pertti Saariluoma

Department of Computer Science and Information Systems, University of Jyväskylä, Finland

P.O. Box 35, FIN-40014 Jyväskylä, Finland; e-mail: pertti.saariluoma@jyu.fi,
tel. +358 14 260 3095, fax +358 14 260 4400

**Corresponding Author*

The effects of alternative navigation device display features on drivers' visual sampling efficiency while searching for points-of-interest were studied in two driving simulation experiments with 40 participants. Given that the number of display items was sufficient, display features that facilitate resumption of visual search following interruptions were expected to lead to more consistent in-vehicle glance durations. As predicted, compared to a grid-style menu, searching information in a list-style menu while driving led to smaller variance in durations of in-vehicle glances, in particular with nine item displays. Kinetic touch screen scrolling induced a greater number of very short in-vehicle glances than scrolling with arrow buttons. The touch screen functionality did not significantly diminish the negative effects of the grid-menu compared to physical controls with list-style menus. The findings suggest that resumability of self-paced in-vehicle visual search tasks could be assessed with the measures of variance of in-vehicle glance duration distributions.

Statement of relevance: The reported research reveals display design factors affecting safety-relevant variability of in-vehicle glance durations and provides a theoretical framework for explaining the effects. The research can have significant methodical value for driver distraction research and practical value for the design and testing of in-vehicle user interfaces.

Keywords: in-vehicle information system; distraction; display; interrupted visual search; resumability; visual sampling strategy

1. Introduction

Solving the problems of human-technology interaction often presupposes deep psychological knowledge. This fact has led to the idea that we should see human-technology interactions as a specific field of modern psychology the way political psychology or traffic psychology is seen (Moran 1981; Oulasvirta & Saariluoma

2004; 2006; Saariluoma & Oulasvirta, in press). Recently, we have called attention to creating explanatory practices, in order to explain and design user interactions with technologies (Saariluoma & Oulasvirta, in press). This means that we must find psychological explanatory models for the observed interaction phenomena, which are consistent with the theories and results of basic psychological research. Drivers' interactions with in-vehicle technologies provide a particularly fruitful framework for applying psychological research to everyday interaction environments.

Driver distraction by in-vehicle information systems (IVISs) is gaining considerable attention due to increasing availability of in-vehicle technologies and services (e.g., Collet, Guillot, & Petit 2010; Drews, Yazdani, Godfrey, Cooper, & Strayer, 2009). These technologies include, for example, factory-installed navigation and entertainment systems, as well as portable hand-held devices, such as mobile phones and music players, which seem to be the most frequent sources of in-vehicle distraction (Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006). The naturalistic 100-car study (Klauer et al., 2006) suggested that visual distraction, in particular, plays a key role in crash and near-crash involvement.

The study of Wierwille, Antin, Dingus, and Hulse (1988) indicated that most of the non-traditional in-vehicle tasks in 1988 required multiple glances away from the driving scene. Since then, the popularity of these types of in-vehicle visual search tasks has substantially increased (Bayly, Young, & Regan, 2008). Drivers can be browsing for visual information among hundreds of similar items in IVIS menus, e.g., a music track, an internet radio station, points-of-interest (POIs), or a person's contact information in phone book contacts. This development places new types of requirements for IVIS display designs.

Simultaneous presentation of numerous items on an in-vehicle display leads to interrupted visual scanning (i.e., visual sampling). If the inspection of all the display items takes over 1.6 seconds of in-vehicle glance time, it becomes uncomfortable to inspect all the items with a single in-vehicle glance due to the building uncertainty of the road events, as suggested by Wierwille (1993). In this case, we propose that the presentation style of the display items becomes a significant factor for the efficiency of visual sampling.

Rapid resumption (RR) phenomenon in interrupted visual search (Lleras, Rensink, & Enns, 2005) indicates that in certain conditions humans are able to resume visual search very efficiently after brief interruptions. However, there is evidence suggesting that additional visuospatial tasks during the interruptions reduce the efficiency of visual search (Shen & Jiang, 2006; Ahn & Lleras, 2007; Ratwani & Trafton, 2008). This is also the case when a driving task and in-vehicle visual search tasks are time-shared. In addition, in these self-paced tasks, the lengths of the interruptions can vary and be significantly longer than the typical interruption intervals in psychological laboratory experiments (e.g., Lleras et al., 2005).

We suggest that by in-vehicle display design we can significantly affect the resumption rates of in-vehicle visual search after interruptions (see e.g., Ratwani, Andrews, McCurry, Trafton, & Peterson, 2007). In particular, familiarity of displays increases systematicity of scan paths in visual search (Rabbitt, 1984). Thus, display features supporting more easily recognizable display item configurations, and thus, points of resumption, should support faster development of efficient, systematic visual scanning strategies. This improved control of resumption of in-vehicle visual search should lead to more controlled in-vehicle glance durations. Accordingly, Victor, Harbluk, and Engström (2005) as well as Hoffman, Lee, McGehee, Macias, and

Gellatly (2005) have noticed that as the demands of an in-vehicle visual search tasks increase, the drivers look at the in-vehicle display for more varied durations.

In this context, the variance, and in particular the tails of in-vehicle glance duration distributions (i.e., the very long glances), have been shown to be more significant for assessing crash risk potential of IVIS use than the average measures of visual demands (e.g., Klauer et al., 2006; Horrey & Wickens, 2007). A study by Horrey and Wickens (2007) points out that the traditional statistical procedures focusing on expected mean values and other measures of central tendency of glance durations can be insufficient for analyzing distraction effects of visual IVISs. By focusing on the average visual demands the safety-relevant tails of glance duration distributions that could reveal the infrequent unintended lapses of control in visual sampling can be left unanalyzed.

In this paper, the measures related to variance of in-vehicle glance duration distributions are called the metrics of visual sampling efficiency. According to Wierwille's (1993) early visual sampling model, efficient in-vehicle glance durations seem to reside between 0.6 to 1.6 seconds. It is assumed that inefficient, unsystematic visual sampling strategies of in-vehicle displays will increase the probability of overlong (over-2-second) in-vehicle glances, which can significantly increase crash risk (Zwahlen, 1988; Klauer et al., 2006).

With two driving simulation experiments, we studied the effects of two alternative display features of a mobile device on drivers' visual sampling efficiency while driving and searching for points-of-interest with mobile navigation software (see Figure 1). These widely-used mobile device features include menu structures of the software (*Grid* or *List*), and menu scrolling methods on a touch screen display

(*Kinetic* or *Buttons*). In addition, the effects of the number of items on a display were analyzed.

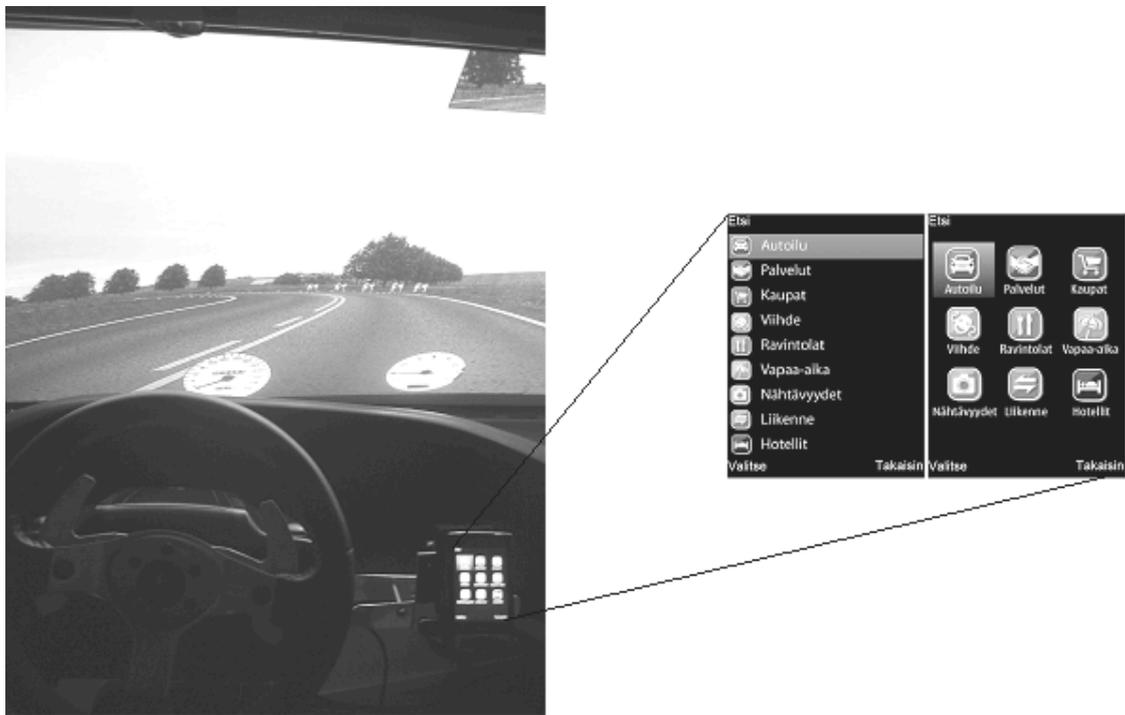


Figure 1. Driving scene from a participant's point of view indicating a leftward lane drift and the two alternative menu structures of the navigation software, *List* and *Grid*.

Compared with the *List*-style presentation of items, a more demanding recognition of the resumption points after interruptions was assumed for the *Grid*-style presentation, where the items appear more similar to each other. In the *List*, items' relative positions and variability in item label length are more easily discriminable than in the *Grid* and provide more efficient discriminative cues for remembering the spatial configuration of the item locations during interruptions (Shen & Jiang, 2006). We have previously observed similar effects with in-vehicle visual search of text types differing on similar aspects (*Spaced* and *Compressed*: Kujala, 2009). Thus, list-wise presentation of items is argued to support more accurate spatial representation of the visited item locations in the spatial memory during interruptions (Ratwani et al., 2007). This should mean better support for systematic, serial inspection of items, with

fewer transitions to already visited items in the *List* menu structure after interruptions (Rabbitt, 1984).

The effects of the menu structure were predicted to become significant at displays of over 6 items, due to the increased requirements for interrupted scanning. Recently, Stevens, Burnett, and Horberry (2010) have shown, with the aid of occlusion measures, that also the calculated visual demand for finding a stock code from a three-column scrolling list is greater than that from a one-column scrolling list.

With the *Kinetic* touch screen scrolling, the continuing variable movement of the items in the menu dependent on the amount of kinetic force applied to the movement was assumed to hinder the recognition of the point of resumption after an interruption. This was supposed to happen in particular when the items still keep moving after driver's gaze has already shifted back to the driving environment. In these cases, the driver's expectations about the contents of the display, i.e., the spatial representation of the item layout configuration (Ratwani et al., 2007), are not met when the driver's gaze returns at the display (Enns & Lleras, 2008). The points of resumption in cases where the *Buttons* support row-by-row scrolling with the up- and down- arrow buttons should be more easily foreseen, enabling a systematic, serial inspection strategy of the new row of items after interruptions.

Our tentative theoretical model suggests that more accurate spatial representations of search item locations during interruptions leads to more systematic visual search during in-vehicle glances (Rabbitt, 1984; Ratwani et al., 2007), which can be observed as more consistent in-vehicle glance durations. Thus, we predicted that the *List* menu structure's and the *Buttons* scrolling method's support for more systematic interrupted visual search than the *Grid* or the *Kinetic* can provide, should become apparent in lower variability of in-vehicle glance durations.

2. Experiment 1 – Effects of menu structure and number of display items on visual sampling efficiency

In *Experiment 1* which featured a mobile device with physical controls, two hypotheses were made based on earlier research. The *List* menu structure was assumed to support more easily resumable serial visual scanning after interruptions than the *Grid* due to the list-wise presentation of items with more easily discriminable text lengths and item positions (Kujala, 2009; Stevens et al., 2010). Thus, based on our tentative theoretical model, we predicted greater variances and skewness in glance duration distributions, greater maximum glance durations, and greater amounts of very long, as well as very short glances towards the display with the *Grid*-style menu. The frequency of very long glances at the display was expected to be higher also while driving in curves, because a lower level of task resumability can affect the driver's ability to time-share visual attention efficiently also in relation to the driving task demands (Lee, Regan, & Young, 2008). Secondly, based on the requirements for interrupted scanning due to the time required to inspect all items at the display (Wierwille, 1993), the effects of the item presentation style were predicted to become significant in particular at displays of over 6 items.

2.1. Method

2.1.1. Design

The experimental design was a mixed-factorial design with 2 x 4 variables (menu structure x number of items). The menu structure (*Grid* or *List*) was a between-subject variable and the number of items in a view (2, 4, 6, or 9) was a within-subject variable. A between-subject design for the menu structure was selected in order to evaluate participants' visual sampling efficiency in their initial exposure to dual-

tasking with the POI search tasks and for enabling comparison between exactly the same search tasks with the different menu structures.

2.1.2. *Participants*

The sample of 20 right-handed participants was randomly selected from the enrolled volunteers recruited via 18 public university e-mail lists. They included 8 women and 12 men ranging in age from 20 to 34 years ($M=25$; $SD=4.0$). All participants had a valid driving license and self-reported lifetime driving experience from 10,000 to 500,000 kilometers ($M=91,000$; $SD=133,000$). Drivers with a very low level of experience as well as aged drivers were not selected for the sample in order to mitigate the known effects of low level of driving experience (Wikman, Nieminen, & Summala, 1998) and aging (Wikman & Summala, 2005) on visual sampling efficiency. All the participants had normal or corrected-to-normal vision. The experiments were conducted, with fluent Finnish-speakers, in Finnish. All the participants were rewarded with a movie ticket.

The participants were divided into two pair-matched groups according to gender, levels of lifetime driving experience (<20,000 km; 20-50,000 km; 50-100,000 km; or >100,000 km) and age (20-25 or 26-35). The group with the *Grid*-style menu had an average lifetime driving experience of 94,000 km ($SD=153,000$), and had an average age of 25 years ($SD=2.8$). For the *List*-group the corresponding averages were 87,000 km ($SD=119,000$) and 25 years ($SD=5.0$).

2.1.3. *Apparatus*

The experiment was conducted in the medium-fidelity, fixed-base three-display driving simulation environment of the Agora User Psychology Laboratory. A recent study by Wang, Mehler, Reimer, Lammers, D'Ambrosio, and Coughlin (2010)

suggests that medium fidelity simulation can provide safe and effective means to evaluate drivers' visual behaviours and task performance with IVIS when compared to on-road studies. The main driving scene was projected into the wind shield of the vehicle cockpit with a resolution of 1280 x 1024 and included a speedometer and a tachometer (see Figure 1). Two side-displays of 22" with a resolution of 1280 x 1024 in the side windows were intended to create a greater feeling of immersion and movement. The motion formulae of the driving simulation software is based on actual engineering documents from the Society of Automobile Engineers (see www.racer.nl). A simulated looped racetrack was used for driving practice, while the actual experiment took place on a simulated rural environment resembling Polish countryside and involving roads with varying curvature. The car simulated was the model 2000 Ford Focus with automatic transmission.

The data collection equipment included consent forms, a SMI iView X HED helmet-mounted eye-tracking system, two video cameras for recording the driving scene with sound and capturing the eye- and scene-videos of the eye-tracking system, and two laptops for capturing the video material. The platform for the navigation software was the Nokia N95 8GB mobile device with 2.8" display positioned in a dashboard holder attached to the right side of the steering wheel column. The distance between the participant and the windscreen-projected driving scene was fixed at 100 cm, but the distance of the pedals and the steering wheel from the participant were adjustable. Hence, the mobile device's distance from the participant varied from 55 to 70 centimeters depending on the length of the participant's arms.

The mobile device had a physical multifunction controller consisting of four-way scroll keys for moving the cursor in the menu and a selection key in the middle for selecting the highlighted item. The two menu structures hold exactly the same

number of items in the different views for each and every menu for every search. The only visible difference between them, in addition to the layout of the items, was the larger icons and the font, which was 1 pt smaller in the *Grid* (font height: 9'-12', icon height: 34'- 44') than in the *List*-menu (font height: 10'-13', icon height: 20'-25', see Figure 1). These small, typical differences in the types of menu structures used were not expected to significantly affect the interpretation of the results. The search targets were delivered aurally, which probably placed an emphasis on a word match (the icons were presumably rarely useful). The *Grid* enabled four-way movements of the cursor in the menu (up-down, right-left), while the *List* supported only two-way movements (up-down).

2.1.4. Procedure

The experiment started with general instructions, signing of a consent form, and adjustment of the pedals and the steering wheel. After the driving rehearsal on a looped track of around 5 minutes, the participant completed a baseline driving task of approximately 10 minutes on a rural road to get more practice and to enable baseline-dual-task comparisons on driving performance. The driving task instructions were to keep the speed of the vehicle between 40-60 km/h, as well as to keep the vehicle in the right lane. There was a speed limit sign of 50 km/h in the beginning of the road to remind the participant about the speed zone. The two Head-Up-Display (HUD) meters between the white lane markings on the road indicated that the vehicle was positioned on the lane. This type of peripheral lane-keeping aid was used in order to enable the participant to focus on the tangent point on the road ahead instead of focusing on the outer edges of the bonnet of the vehicle. In addition, an immediate full stop was instructed in the event of the participant seeing a deer somewhere in the environment.

Driving practice included the deer reaction task, while the baseline or dual-task driving did not. The participants were not made aware of this beforehand. The participants encountered oncoming traffic in the form of four cars at preset points on the road, but were not required to interact with them. The other cars were included and the deer observation task was instructed in order to stress that there was the possibility of unexpected events and to make the participants observe the environment in a more natural way, rather than merely observing the lane markings and the speedometer.

Before the dual-task drive, the participants got to practice the search task once without the driving task on the *Grid-* or *List-* style menu. For the dual-task drive on the rural road, the participants were instructed to keep their *priority on driving* and the search tasks were *self-paced*. Driving task priority in the dual-task driving was further emphasized by a promise of an additional movie ticket to the 10 most accurate drivers. Driving task accuracy was defined as the total time the HUD meters were positioned out of the lane (for disambiguating where the vehicle's edges were) and the speed was above or below the instructed speed zone. A hypothetical scenario was used where the participants were asked to imagine that they were travelling in Poland by car and searching for points-of-interest located nearby. Table 1 represents the search tasks that were given to the participants in a randomized order, the pair-matched participants having the same task orders.

[Insert Table 1 about here]

The number of items in a view varied from 2 to 9 within the tasks depending on which menu in the sequence was being displayed (see Figure 2). In tasks 6, 7, and 8, the participant had to scroll the last 9 item-per-view menus. The push of a down scroll key revealed one row (or item: *List*) at a time once the cursor was in the lowest

row at the display. The experimenter gave the tasks verbally during the driving, allowing for a very short pause of a few seconds between tasks after a successful task. Task instructions could be repeated by saying “repeat”, if the participant forgot or did not hear the task. The first task was initiated when the participant reached 40 km/h for the first time. Search tasks started at random points on the road depending on the participant’s performance. In the dual-task condition, every other participant drove the same road as the rest, but in the opposite direction, which kept the driving task demands (i.e., road curvature) at the same level for every participant but provided more randomness to the task starting points.



Figure 2. Task 7: ‘Find the way to a theatre named Kto’ with the *List* (upper row) and the *Grid* (lower row) menu structures. The right selections are illustrated with the white rectangles. The last view required scrolling the menu in order to find Kto.

The dual-task driving lasted from 6 to 10 minutes depending on the participant’s individual performance. After driving, the participant was interviewed in order to explore the participant’s search strategies, task prioritization, and to classify the drivers’ ways of interacting. The main questions of interest were:

- “How did you perform the search tasks? Did you have some search strategy or were you able to develop a search strategy during the drive?”, and
- “Could you imagine yourself conducting this type of search activity while driving?”.

2.1.5. Variables and analysis

The independent variables were the menu structure and the number of items in view. The principal dependent variables measured the efficiency of visual sampling, but also visual demands, driving performance, and search task performance were assessed in order to see the relationship of visual sampling efficiency with these variables.

Visual sampling efficiency was measured by the maximum and standard deviations of glance durations (at the mobile device), by the frequency of over-1.6-second and over-2.0-second in-vehicle glances in total and while driving in curves, as well as by the frequency of under-0.4-second glances. The shapes of the glance duration distributions were analyzed by the measure of skewness, i.e., the asymmetry of the probability distribution of a real-valued random variable. The upper limit of 1.6 seconds was selected because it has been observed that drivers generally prefer to keep their glances at in-vehicle displays below this duration in most circumstances (Wierwille, 1993; see also Wang et al., 2010). Over-2.0-second glances can be associated with higher crash risks and the frequency of near crash situations as well as minor incidents in real traffic (Klauer et al., 2006). The frequency of these overlong glances while driving in curves served the purpose of assessing drivers’ ability to consider the demands of the driving situation in time-sharing visual attention between tasks. The movement of gaze from the driving scene to the device ($M=160$ ms) and back was scored into the glance duration (SAE, 2000). The frequency of the very short, under-0.4-second glances was taken as an indicator of uncertainty regarding the

status of the display (see Hoffman et al., 2005; Wikman et al., 1998). The effects of the number of items were analyzed for maximum and standard deviations of glance durations in the cases where there were enough glances to enable meaningful analysis. In addition, interaction effects of menu structure and the number of items on these measures were analyzed.

Total number and average duration as well as total duration of glances (total glance time, *TGT*) at the device served as measures of visual demands. These are the most often used measures for assessing the visual demands of secondary tasks (Green, 1999). Driving performance was measured as the total number (*NLE*) and duration (*DLE*) of lane excursions and speed maintenance errors (number: *NSE* and duration: *DSE*). Search task performance was measured as the frequency of errors, defined as a selection of a wrong item, and task completion times with driving excluded, i.e., total glance times at the display by task. The within-subject effects of the number of items on driving performance were excluded in the analysis of driving performance and search task performance due to the difficulty of the scoring process.

Search task performance, lane excursions, and eye-movements were scored manually frame-by-frame (25 frames per second) with the Noldus Observer XT software. A single glance at the in-vehicle display was scored, following the SAE J2396 definition (SAE, 2000), from the video images provided by the eye-tracking system, indicating participants' eye movements in the eye-video and head movements in the scene-video. This scoring method meant more work and made it unfeasible to analyze accurate fixation data, but on the other hand, assured that there was no data loss due to possible technical faults. Linking the number of items to the glances was done by noting the number of items visible on the display at the moment the glance was scored to start. An automatic script compared the steering wheel movements

recorded in the log file of the driving simulation to the synchronized eye-tracking data file for scoring the frequencies of overlong glances in curves. The absolute value of 1.00 or more of the steering wheel position in terms of the simulation's log file data was selected as the limit for driving in a curve. The value of 0.00 indicated the calibrated centre position of the steering wheel. The number and durations of lane excursions were analyzed for equal journey lengths between the baseline and dual-task driving. The total number and duration of speed maintenance errors were scored automatically with a script from the simulation log file, excluding the first acceleration and last deceleration phases.

The simulated road had an end, meaning that there was a time limit of about 10 minutes for the completion of the search tasks. Three of the participants did not have enough time to start or complete the last tasks in their drives. This was balanced in the analysis by excluding the corresponding task data from their pairs in the other group. The participants were not told about the time limit.

Two-tailed *t*-tests were utilized for between-subject comparisons on visual sampling efficiency, visual demands, and search task performance. Repeated measures ANOVA was used in order to find interaction effects on driving performance (menu x baseline/dual-task condition, 2 x 2) and to account for differences in baseline performance. ANOVA was used also for those variables for which the scoring process enabled the effects of the number of items to be assessed (menu x items, 2 x 4). A .05 alpha level was used in the statistical testing. The interviews were analyzed from the videos and the frequencies of yes/no answers were calculated.

2.2. Results and discussion

2.2.1. Visual sampling efficiency

As hypothesized, the maximum glance durations $t(18)=3.03$, $p=.007$, as well as the frequencies of over-1.6-second, $t(18)=2.64$, $p=.017$, and over-2.0-second, $t(18)=2.21$, $p=.040$, glances in total and while driving in curves (over-1.6-second: $t(18)=2.24$, $p=.038$; over-2.0-second: $t(18)=2.53$, $p=.021$) were significantly greater for the participants using the *Grid* than for the participants using the *List* menu structure for the in-vehicle search tasks (see Figure 3).

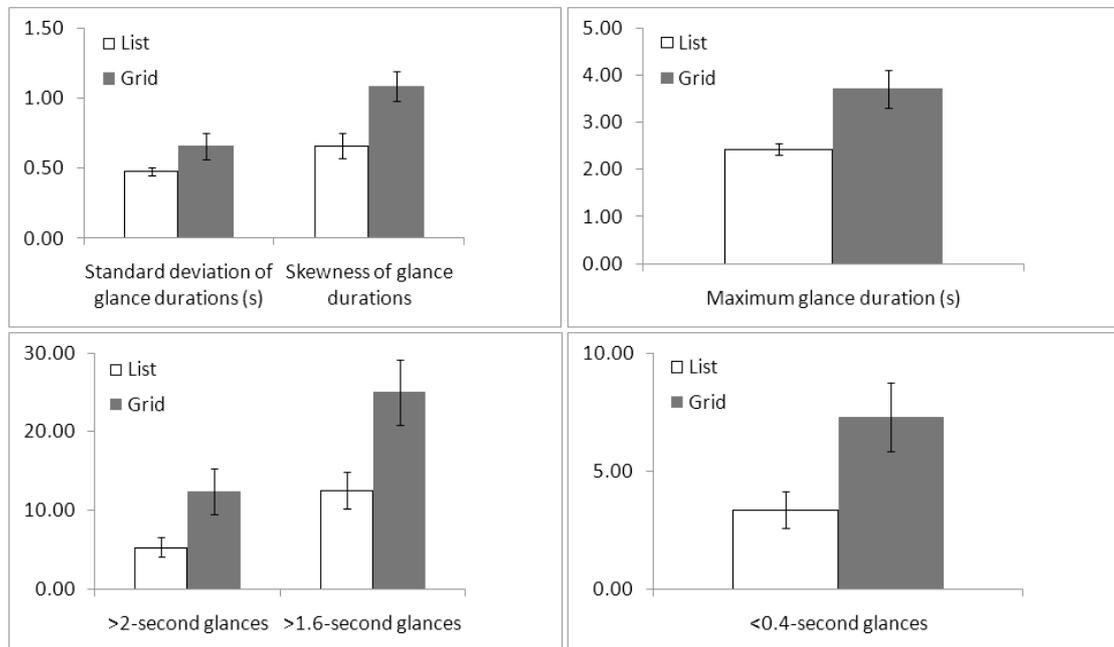


Figure 3. Effects of menu structure on the metrics of visual sampling efficiency in *Experiment 1*. Means and SEs.

In addition, there was a significantly greater number of under-0.4-second glances towards the display with the *Grid* menu, $t(18)=2.39$, $p=.028$. Considering the variances in glance duration distributions, the standard deviation of glance durations only approached significance with these sample sizes, $t(18)=1.96$, $p=.065$, but the skewness of glance duration distributions was significantly lower for the *List* than for the *Grid*, $t(18)=2.97$, $p=.008$. As predicted, the number of items on the display had a

significant increasing effect on the maximum glance durations, $F(1,18)=46.69$, $p<.001$ (see Figure 4).

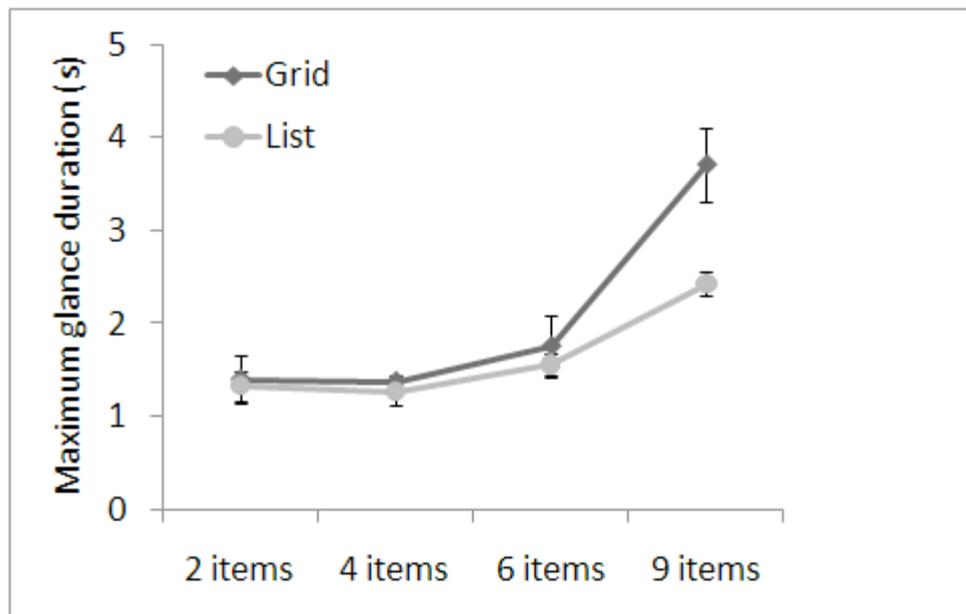


Figure 4. Maximum glance durations by menu structure and the number of items in *Experiment 1* ($n=10$). Means and SEs.

2.2.2. Visual demands

The participants had significantly lower total glance times, $t(18)=3.04$, $p=.007$, and number of glances, $t(18)=2.45$, $p=.025$, at the device in tasks with the *List* menu (see Table 2). This finding is in line with the study of Stevens et al. (2010) indicating greater calculated visual demands for finding a stock code from a three-column compared to one-column scrolling list. Average glance durations did not differ significantly between the menu structures. This finding supports the conclusions of Horrey and Wickens (2007) that the average in-vehicle glance durations can be at a similar level between IVIS designs, although there could be safety-critical differences in the glance duration distributions.

2.2.3. Driving performance

Overall, even if approaching significance, the dual-task condition did not have a significant effect on the number or duration of lane excursions or speed maintenance errors ($F(1,18)=3.35$, $p=0.084$). There were no significant between-subject effects of the menu structure or interaction effects of the dual-task condition and the menu structure despite the observed effects on visual behaviours. Lennemann and Backs (2009) as well as Wang et al. (2010) have shown evidence that the absence of dual-task costs in driving performance does not have to mean that there are no attentional costs of dual-tasking while driving.

2.2.4. Search task performance

The participants completed Tasks 2 (*List*: 58.25 (1.17), *Grid*: 141.55 (3.95), $t(16)=2.25$, $p=.039$) and 6 (*List*: 123.60 (.81), *Grid*: 241.22 (5.25), $t(16)=2.46$, $p=.026$) significantly faster with the *List* menu than with the *Grid* menu. There is no obvious explanation for why there were differences only between these two tasks, but this could relate to errors made in these two tasks (Task 2 (SUM): *List*: 1, *Grid*: 6, Task 6 (SUM): *List*: 0, *Grid*: 5). In total, there was a significantly greater number of wrong selections in the tasks with the *Grid*-style menu ($M=5.9$, $SE=1.2$) compared to the *List*-style menu ($M=2.7$, $SE=.4$), $t(18)=2.49$, $p=.023$.

During the dual-tasking, the participants learned to find and select the often repeated functions of the software (e.g., Options and Search) without visual attention. A possible alternative explanation for the more efficient visual sampling with the *List*-menu, besides the items' spatial configuration, as well as for the lower total glance durations and lower number of glances relates to this tactical finding. With *List*, the participants were often able, after locating the target item, to quickly estimate the required steps to the item and perform the movement of the cursor without visual

attention. The *Grid*-style menu did not seem to support this. When the participants were asked about their willingness to engage in this type of search activity while driving, 17 out of the 20 participants reported this as highly feasible. The three unwilling participants were in the *Grid* group.

3. Experiment 2 – Effects of scrolling method and touch screen functionality on visual sampling efficiency

Experiment 2 featured a mobile device with a touch screen display and *Grid*-style menus with two typical ways of scrolling menus on a touch screen. It was assumed that touch screen scrolling with the arrow buttons would support more systematic interrupted visual scanning with a higher level of resumability than the kinetic scrolling method. This was intuitively the case, because kinetic scrolling of menu by fingertip induces more variation and instability in the ways the menu brings up more items. In particular, if the menu is scrolled with a kinetic force, the items can still keep moving after the driver's gaze has already shifted to the driving environment. This leads to incorrect expectations of the spatial configuration of the display items when the gaze is returned at the display (see Ratwani et al., 2007). The upcoming points of resumption would be more easily foreseen with the method supporting row-by-row scrolling by up- and down- arrow buttons. This was expected to be reflected by greater variances in glance duration distributions with the kinetic scrolling. Again, an increasing number of items at the display was expected to decrease visual sampling efficiency with the *Grid*-style menus.

The tactical findings in *Experiment 1* related to the cursor movements brought up new questions. There was a difference in the complexity of the required cursor movements between the *List* (two-way movements) and *Grid* (four-way movements). This meant that the *Grid* did not support as well as well as the *List* the tactical

behaviour of not looking at the display while moving the cursor. The comparison of the data of *Experiment 1* with the data of *Experiment 2* collected from the same tasks but with a touch screen device with the *Grid*-menus, should reveal whether the inefficient visual sampling with the *Grid* menu was due to the more complex movements, with physical controls, of the cursor in the menu, or to the spatial configuration of the display items. We expected support for the latter explanation as significant differences in visual sampling efficiency between the *Grid*-menus on the touch screen device and the *List*-menus on a device with the physical scroll keys, even though the touch screen eliminates the complex cursor movements.

3.1. Method

3.1.1. Design and procedure

Again, the experimental design was a mixed-factorial design with 2 x 4 variables (touch screen scrolling method x number of items). The procedure was exactly the same as in *Experiment 1*, but with additional instructions and practice on the scrolling method. In addition, the two experiments presented in this paper were conducted partly at overlapping dates.

3.1.2. Participants

A sample of 20 right-handed volunteers, 8 women and 12 men, with ages ranging from 20 to 35 ($M=26$; $SD=3.5$), were selected randomly from the enrolled volunteers but divided into two groups in order to form comparable groups with the groups of participants of *Experiment 1*. The participants had not taken part into *Experiment 1*. Participants' self-reported lifetime driving experience varied from 25,000 to 400,000 kilometers ($M=100,000$; $SD=89,000$). Again, all the participants had normal or

corrected-to-normal vision, and the experiments were conducted in Finnish with fluent Finnish-speakers.

In a similar manner to *Experiment 1*, the participants were divided into two pair-matched groups. The group members with the kinetic-style scrolling method had an average lifetime driving experience of 93,000 km ($SD=67,000$), and an average age of 26 years ($SD=3.5$). For the group with the buttons-style scrolling method the corresponding averages were 108,000 km ($SD=110,000$) and 27 years ($SD=3.6$).

3.1.3. Apparatus

Exactly the same driving simulation environment, simulated roads, simulated Ford Focus, and the same data collection systems as in *Experiment 1* were used in *Experiment 2*. The tasks, scenario, and basic instructions for the participants were identical to *Experiment 1*. The only difference between the experiments was the mobile device. This time the device used was Nokia XpressMusic 5800 with 3.2” touch screen. The navigation software was the same as in *Experiment 1* but modified for touch screen use.

The menu structure held exactly the same number of items in the different views for each and every menu for every search task as the menu structures of *Experiment 1* did. The size of the icons and fonts were identical to those of the *Grid* menu structure in *Experiment 1*. However, the items on the display were selectable with touch, and there were two different methods for scrolling menus featuring over 9 items in the Tasks 6, 7, and 8: arrow buttons and kinetic scrolling. The two arrow buttons (up arrow and down arrow) revealed one row and hid one row of three items in the menu at the touch of a button on the touch screen, in an identical fashion to that of the physical scroll keys in *Experiment 1*. Kinetic scrolling here meant that the participant could reveal more rows in the menu with a fingertip by gently pressing the

touch screen and by moving the finger in the desired direction (up or down). The movement of the items depended on the speed of the finger movement. The menu kept scrolling after the finger movement relative to the 'kinetic force' applied to the movement. This feature differentiates the kinetic-style scrolling from 'sweeping', which consists of a well-defined movement of the menu with a single finger displacement on a touch screen. Both scrolling methods revealed a scroll bar indicating the current position of the menu when a button was pressed or the menu scrolled with the kinetic style, otherwise the scroll bar remained invisible.

3.1.4. Variables and analysis

The independent variables were the touch screen scrolling method and the number of items in a view. Also the effects of the number of items were analyzed. The dependent measures included, once more, visual sampling efficiency, visual demands, driving performance, and search task performance. The metrics for assessing these variables were the same as in *Experiment 1*. The mitigation of the undesired effects of the controlled variables and scoring of the behaviours were conducted in exactly the same manner.

Finally, the data of the current experiment was juxtaposed with that of *Experiment 1*: comparisons were made between the touch screen device and the device with physical controls, as well as between the effects of the other user interface features. In addition, correlations between the metrics were calculated using the Pearson's *R* correlation coefficient with two-tailed tests of significance ($N=40$) in order to assess possible statistical relationships. For enabling balanced data, the corresponding data of the pair-matched participants in *Experiment 2* was excluded from the analysis where data was missing in *Experiment 1*. Again, two-tailed *t*-tests were utilized for between-subject comparisons and repeated-measures ANOVA was

used for the analysis of the significance of interaction effects. One-way and repeated-measures (for the analysis of driving performance) ANOVAs with the Bonferroni correction were utilized for multiple comparisons. The alpha level was set to .05. The interviews were conducted and analyzed in the same manner as in *Experiment 1*.

3.2. Results and discussion

3.2.1. Visual sampling efficiency

The only significant effect of the scrolling method was found on the lower frequency of under-0.4-second glances with the *Buttons* scrolling method, $t(18)=2.72, p=.014$ (see Table 2). Maximum glance durations, $t(18)=1.92, p=.071$, standard deviation of glance durations, $t(18)=1.70, p=.106$, or the skewness of glance duration distributions, $t(18)=1.81, p=.086$, although indicative, did not indicate significant effects of the scrolling method. With more statistical power these effects could have been significant. Again, the number of items on the display had a significant increasing effect on maximum glance durations (two items: $M=1.64, SE=.13$; four items: $M=1.56, SE=.09$; six items: $M=1.94, SE=.19$; nine items: $M=3.20, SE=.19$), $F(1,18)=33.35, p<.001$. This gives further support to the findings of *Experiment 1* on the significance of the number of items at the display. A closer analysis reveals that *Kinetic* scrolling ($M=9.9, SE=1.8$) induced a significantly greater number of under-0.4-second glances at 9-item views than *Buttons* ($M=2.3, SE=.8; t(18)=3.84, p=.001$) did.

[Insert Table 2 about here]

3.2.2. Visual demands

There were no significant effects of the scrolling method on total glance times, total number of glances, or average glance durations.

3.2.3. *Driving performance*

Overall, the dual-task condition had a significant effect on the number, $F(1,18)=10.48, p=.005$, and duration, $F(1,18)=7.40, p=.014$, of lane excursions but not on speed management (see Table 2). However, in the dual-task drive the group with the button scrolling method made significantly fewer speed management errors than the group with the kinetic scrolling method, $F(1,18)=6.88, p=.017$. A more detailed analysis revealed that the participants with the *Kinetic* scrolling method made significantly more speeding errors (mean difference (*Kinetic-Buttons*)=1.4(.4), $F(1,18)=9.31, p=0.007, 95\%CI=.42$ to 2.28) with longer total durations (mean difference (*Kinetic-Buttons*)=10.11(4.64), $F(1,18)=4.75, p=0.043, 95\%CI=.37$ to 19.85) than the participants with the *Button* scrolling.

In addition to inefficiency in task timing, poorly designed visual in-vehicle tasks can lead to increased levels of working memory load, which has been observed to affect particular reaction times (e.g., Jahn, Oehme, Krems, & Gelau, 2005) and speed management (e.g., Recarte & Nunes, 2002). This could possibly explain the higher number of speed management errors with *Kinetic* scrolling compared to *Buttons*.

3.2.4. *Search task performance*

The only significant effect of the scrolling method on search task performance was that the participants performed Task 8 (27 items in the last menu) significantly slower (*TGT*) with the kinetic scrolling method ($M=46.82, SE=9.18$) than with the arrow buttons ($M=24.56, SE=3.73$), $t(18)=2.25, p=.039$.

It should be noted that neither of the tested scrolling methods proved to be optimal.

The scroll bar was visible only for a short duration during scrolling, and both methods

required accuracy for hitting the correct arrow button or for positioning the finger between the items on the display. In addition, some of the participants reported that it was not at first clear how many items were revealed with one push of an arrow button. Moreover, Hoffman et al. (2005) have shown that line-by-line scrolling (of text) with a touch screen button can place more visual demands on the user than page-by-page scrolling due to the increased requirement to locate and use the scroll button. Page-by-page scrolling could be a better option with touch screen buttons, but presumably with a list-wise presentation of the items. In addition, the touch screen provided no sound or tactile feedback for selecting the items. However, 14 out of the 20 participants reported that they could imagine themselves conducting these types of search tasks while driving.

3.2.5. *Multiple comparisons and correlations*

The multiple comparisons of the data of *Experiment 1* and *Experiment 2* with the Bonferroni correction indicate that there were significant differences between the effects of the different features (see Table 2). Touch screen functionality did not seem to support as efficient interaction as the physical controls with the *List* menu for either the group with *Buttons* scrolling (*TGT*: mean difference (*Buttons-List*)=68.52(17.86), $p=.003$, 95%CI=18.65 to 118.38; total number of glances: mean difference (*Buttons-List*)=48.7(15.7), $p=.022$, 95%CI=4.9 to 92.5) or the group with *Kinetic* scrolling (maximum glance duration: mean difference (*Kinetic-List*)=1.18(0.40), $p=.033$, 95%CI=.07 to 2.29; under-0.4-s glances: mean difference (*Kinetic-List*)=6.2(2.0), $p=.021$, 95%CI=.66 to 11.74).

However, the worst visual time-sharing performance was still observed in the case of the *Grid* menu with the physical controls when comparing the number of over-2.0-second glances in curves (mean difference (*Grid-Buttons*)=3.2(1.1), $p=.033$,

95%CI=.18 to 6.22; mean difference (*Grid-Kinetic*)=3.4(1.1), $p=0.20$, 95%CI=.38 to 6.42). Figure 5 indicates the effects of the touch screen scrolling method (*Kinetic* or *Buttons*) and menu structure (*List* or *Grid*, in a device with physical controls), illustrating participants' visual sampling efficiency with these tasks.

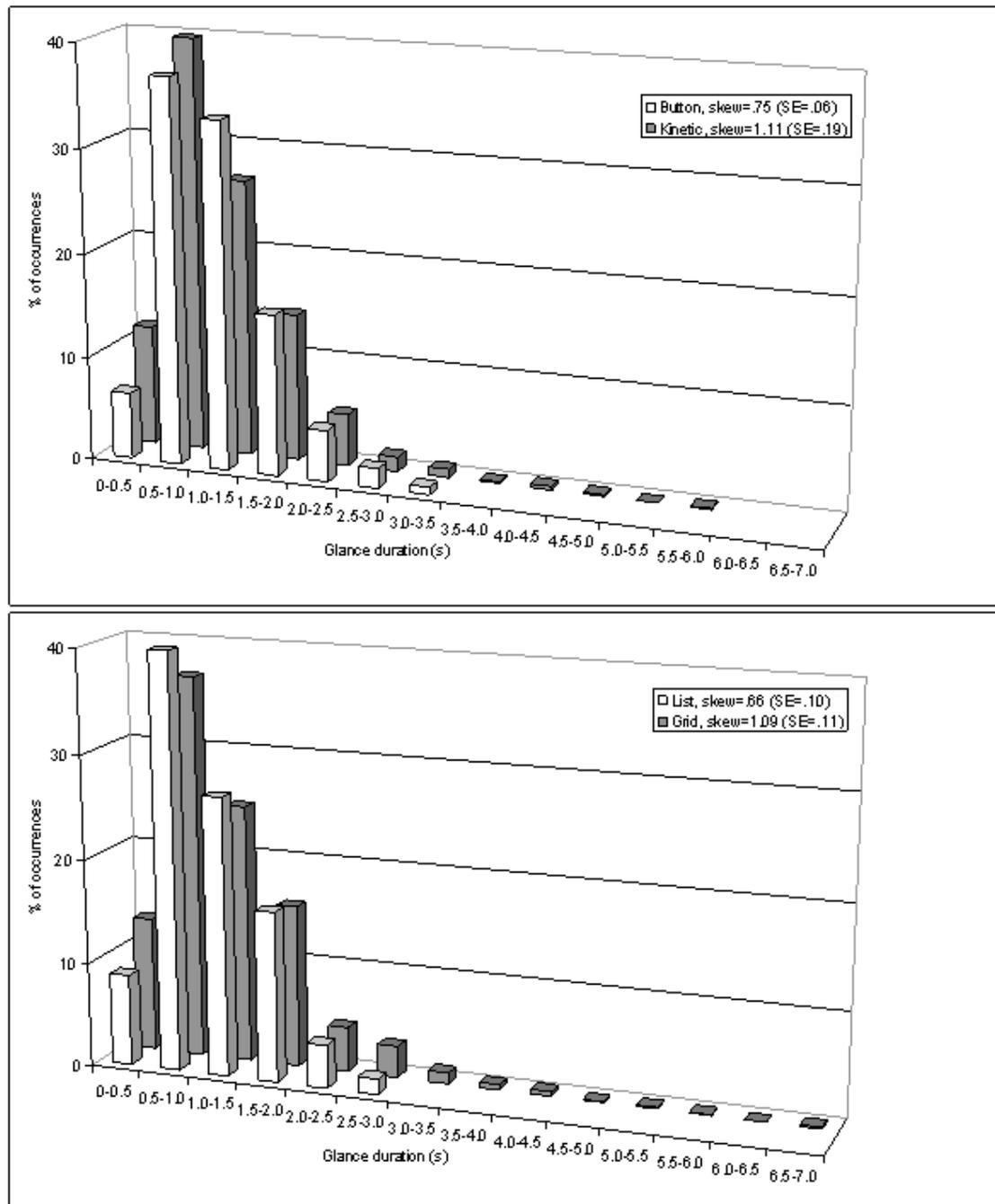


Figure 5. Effects of the scrolling method (*Kinetic* or *Buttons*) in a touch screen device and the menu structure (*List* or *Grid*) in a device with physical controls on participants' glance duration distributions.

The total data (*Exp 1* and *Exp 2*, $N=40$) indicated significant positive associations between skewness of the individual glance duration distributions, with maximum glance durations, $r=.60$, $p<.001$, standard deviation of glance durations, $r=.42$, $p=.008$, and frequency of glances shorter than 400 ms, $r=.45$, $p=.003$. Glance duration distributions at in-vehicle devices typically show a positively skewed distribution toward short glances (Victor, Engström, & Harbluk, 2008). It seems that the measure of skewness is more sensitive than standard deviation (*Experiment 1*) and the most safety-relevant variable for assessing the variance of the log-normally distributed glance duration distributions. This is due to its relation to the extent of the right-hand tail of the distribution. The skewness of the distribution is also increased by more mass concentrated on the far left side of the mode duration, i.e., by the very brief glances.

The significant associations between lane excursion and eye-tracking measures are listed in Table 3. The speed maintenance metrics had significant associations with total glance times (NSE: $r=.32$, $p=.041$; DSE: $r=.51$, $p=.001$) and total number of glances (NSE: $r=.33$, $p=.037$; DSE: $r=.48$, $p=.002$).

[Insert Table 3 about here]

4. General discussion

The purpose of the current experiments was to analyze the effects of different menu structures and touch screen scrolling methods of a mobile device on drivers' visual sampling efficiency while driving and searching for points-of-interest with mobile navigation software. As predicted, the results indicate that the participants' first-time interaction with the *List*-style menu structure while driving led to smaller variance in glance durations at the display than with the *Grid*-style menu. This effect

was noteworthy in particular on maximum glance durations at displays featuring nine items per view.

The effect of the number of display items can be explained with the in-vehicle glance time required to inspect all the items (Wierwille, 1993), which leads to interrupted visual search. An additional explanation is the increased difficulty to keep in mind the point of resumption and the locations of the inspected items during interruptions due to visual short-term memory capacity limitations (4+-2 items: Cowan, 2005) when a sufficient mental model of the display contents has not yet been formed. Visual search that is related to the driving task with spatial elements can also interfere with the efficient coding of the spatial display layout (Shen & Jiang, 2006; Ratwani & Trafton, 2008).

Although not observed in average glance durations, the effects of menu structure became visible also in total glance times and total number of glances. While drivers seem to be generally aware of the risks of visual distraction and try to keep in-vehicle glance lengths below 1.6 seconds (Wierwille, 1993), the increased visual demands of in-vehicle tasks can lead to occasional failures in the control of visual sampling as suggested by Horrey and Wickens (2007).

As expected, the results of *Experiment 2* indicated that on a mobile device with a touch screen display, the menu scrolling method with arrow buttons supports lower variability in glance durations than the kinetic scrolling method, in particular when comparing the number of very short (<400 ms) glances. Although not safety-critical or necessarily ineffective as such, the very short in-vehicle glances can indicate increased requirements to quickly check the status of the display due to uncertainty. An increased number of these glances was observed already with the *Grid* menus compared to the *List* menus in *Experiment 1*. In *Experiment 2*, the finding

may relate to the occasional brief glances the drivers made to check when (or where) the kinetic scrolling of the display items had stopped (for similar findings, see Hoffman et al., 2005).

The observed differences in the variability of in-vehicle glance durations due to IVIS display features can be explained with the features' support for the accuracy of drivers' spatial representations of item locations during interruptions and thus, resumability of scanning. Higher accuracy of these spatial representations means better support for controlled and systematic serial visual search strategies (Rabbitt, 1984). If the driver has insufficient or no mental model of the locations of the inspected items at the display, such factors as saliency of the items or some general scanning strategy (Underwood, Foulsham, & Humphrey, 2009) can lead the gaze to already inspected items after interruptions. This means inefficient use of in-vehicle glance time and a possible time cost. Uncertainty regarding the contents of a display can increase the lengths and irregularity of scan paths at the display (Rabbitt, 1984; Goldberg & Kotval, 1999). Uncertainty can also increase the attention-capturing effects of the salient items at the display (Foulsham & Underwood, 2009; Underwood et al., 2009), fixation durations due to unfamiliar or unexpected items (Friedman, 1979; Underwood & Everatt, 1992), and the frequency of very short glances (Wikman et al., 1998; Hoffman et al. 2005). All these effects lead to more random, varied glance durations at the display compared to more controlled interrupted visual search.

The touch screen functionality did not significantly diminish the negative effects of the *Grid*-style menu on visual sampling efficiency compared to the physical controls with the *List*-style menu. This seems to indicate that these effects are mainly due to the visual layout of the *Grid*-style menu, as expected. However, the more complex four-way movements of the cursor required with the physical controls in the

Grid menu seemed to further increase the frequency of over-2-second glances at the display while driving in curves. A possible explanation could be that because it is demanding to keep in mind the location of and the path to the target item in the *Grid* menu, the driver is encouraged to make the selection during the on-going glance immediately after locating the target item. This is even when the selection would have negative effects on in-vehicle glance times and even lane-keeping. Brumby, Salvucci, and Howes (2009) have shown that drivers are willing to sacrifice some lane-keeping accuracy in order to overcome cognitive constraints of visual secondary tasks. With the *List* menu, the participants were often able to make the required cursor movements without the aid of visual attention. Naturally, the touch screen functionality disabled all tactical abilities for selecting items without visual attention.

Considering the driving performance effects of touch screen use, there was a significant negative effect of dual-task condition on lane-keeping in *Experiment 2* with the touch screen and the *Grid*-menus, but no such effect in *Experiment 1* with the physical controls and the *List* or *Grid* menus. However, this finding can be attributed to the use of only *Grid*-style menus in *Experiment 2*. Furthermore, the attentional costs of visual in-vehicle tasks do not necessarily have to have any direct bearing on decreases in driving performance (see Lennemann & Backs, 2009 and Wang et al., 2010), although these can appear as a significant contributor to crash risk due to the increased probability of overlong glance lengths at the in-vehicle devices in response to the demands of the driving situation. Overall, in our experiments the increased number of over-1.6-second and over-2-second glances, in particular while driving in curves, were highly associated with reduced lane-keeping performance.

The greater number of secondary task errors with the *Grid* menu structure compared to the *List* gives further support to our assumption that visual search in the

Grid menus is more complex when combined with driving. Neither of the inefficient features – i.e., the *Grid* menu structure or the *Kinetic* scrolling method– seem to be complex to use while one is allowed to concentrate on the search task alone.

However, when the use of them is combined with the driving task, the level of task complexity increases due to interruptions, demands to memorize the locations of the items at the display, and the interfering driving task demands for visual-spatial processing. This is the reason why bench-testing or the occlusion method (ISO, 2007) may not be sufficient (see also Monk & Kidd, 2007). IVIS display designs should be tested also with self-paced actual or simulated driving with the priority in the driving task.

According to Rockwell (1988), drivers try to develop consistent visual sampling strategies with visual in-vehicle tasks. Individual learning through practising a task increases the level of consistency and automaticity in visual search behaviours (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977; Rabbitt, 1984). Research on expertise and automatization suggests that the more extensive the task-related skills and the domain-specific knowledge, the lower the variance in task-relevant information-seeking behaviours and scan patterns (Underwood et al., 2009) as well as in the resultant task performance (e.g., Logan, 1988; Rasmussen, 1990). Thus, the metrics of visual sampling efficiency can be used for assessing how easily an efficient visual sampling strategy can be learned and automaticity in in-vehicle visual search patterns developed. The metrics can also be used in the development of distraction warning systems for noticing inefficient visual sampling behaviours and the related potential for visual distraction before it has negative consequences for the operational control of the vehicle.

Our research indicates that as the number of items on in-vehicle displays presented to the driver increases, the issues of presentation style and ways of browsing the items become increasingly important. Our results suggest that to avoid the risks related to visual sampling efficiency due to display item presentation styles, the maximum number of items simultaneously at the display should be limited to six. However, the *List*-style presentation of the items could allow a safe presentation of even nine items at once. The possibility to break the in-vehicle search task into clear, systematic, and easily resumable steps of visual search, with discriminative cues for search items' spatial configuration aiding in the resumption of the scanning task, should be a basic requirement for all in-vehicle visual search tasks. The *Grid* menu structure and the *Kinetic* scrolling method do not seem to support this as well as the *List* menu structure and the *Button* scrolling. Similar examples of presumably risky IVIS user interface features requiring testing include, for example, stepless zoom and pan features with fingertip gestures, common in several modern touch screen devices. The negative effects of the display features on visual sampling efficiency are naturally likely to be magnified among elderly (Wikman & Summala, 2005) or novice drivers (Wikman et al., 1998).

Future research should analyze, in detail, fixation durations, scan paths and resumption rates on in-vehicle displays in order to provide further support for our tentative theoretical considerations on the relationship between systematic scanning patterns and consistency of in-vehicle glance durations. Further research should address, among a vast amount of other display design issues, more carefully the combined effects of *List*-style menus and a touch screen display compared to the observed effects of *Grid*-style menus. The current partly confounded experimental designs, although indicative, did not enable detailed analyses on whether the observed

differences were due to factors related to discriminative cues, cursor movement, or menu scrolling and thus, could not shed light on these factors' relative importance for visual sampling efficiency. Nonetheless, it is important to see that our results are consistent with the general research on interrupted visual search, skilled behaviour, and automatization. The research seems to support our conclusions and enabled us to explain the found interaction phenomena based on general properties of human attention.

Acknowledgments

The authors want to express their gratitude especially to Mikko Nirhamo at Nokia for his valuable collaboration. We also want to thank the two anonymous reviewers for providing helpful and constructive comments and thus helping us to significantly improve the quality of the paper. This work was supported by Theseus, a research collaboration project on human-technology interaction funded by TEKES (Finnish Funding Agency for Technology and Innovation). The driving simulation environment was developed by a grant from the Henry Ford Foundation.

References

- Ahn, J.-W. & Lleras, A. (2007). Spatial working memory loads can reduce search efficiency but not the rates of rapid resumption in interrupted visual search tasks. *Journal of Vision*, 7(9), article 685.
- Bayly, M., Young, K.L., & Regan, M.A. (2008). Sources of distraction inside the vehicle and their effects on driving performance. In M.A. Regan, J.D. Lee, and K.L. Young (Eds.), *Driver Distraction: Theory, Effects, and Mitigation* (pp. 191-214). Boca Raton, FL: CRC Press.
- Brumby, D.P., Salvucci, D.D., & Howes, A. (2009). Focus on driving: How cognitive constraints shape the adaptation of strategy when dialing while driving. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI'09* (pp. 1629-1638). New York, NY: ACM Press.
- Collet, C., Guillot, A., & Petit, C. (2010). Phoning while driving I: a review of epidemiological, psychological, behavioural and physiological studies. *Ergonomics*, 53, 2010, 589-601.
- Cowan, N. (2005). *Working Memory Capacity*. New York, NY: Psychology Press.
- Drews, F.A., Yazdani, H., Godfrey, C.N., Cooper, J.M., & Strayer, D.L. (2009). Text messaging during simulated driving. *Human Factors*, published online 16 December 2009: <http://hfs.sagepub.com/content/early/2009/12/16/0018720809353319>.
- Enns, J.T. & Lleras, A. (2008). What's next? New evidence for prediction in human vision. *Trends in Cognitive Sciences*, 12(9), 327-333.
- Foulsham, T. & Underwood, G. (2009). Does conspicuity enhance distraction? Saliency and eye landing position when searching for objects. *The Quarterly Journal of Experimental Psychology*, 62, 1088-1098.

- Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, *108*, 316-355.
- Goldberg, J.H. & Kotval, X.P. (1999). Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics*, *24*, 631-645.
- Green, P. (1999). *Visual and Task Demands of Driver Information Systems*. Report No. UMTRII98-16. Ann Arbor, MI: University of Michigan Transportation Research Institute.
- Hoffman, J.D., Lee, J.D., McGehee, D.V., Macias, M., & Gellatly, A.W. (2005). Visual sampling of in-vehicle text messages: Effects of number of lines, page presentation, and message control. *Transportation Research Record*, *1937*, 22-30.
- Horrey, W. J. & Wickens, C. D. (2007). In-vehicle glance duration: Distributions, tails, and model of crash risk. *Transportation Research Record*, *2018*, 22-28.
- International Organization for Standardization (2007). *ISO 16673 Road Vehicles- Ergonomic Aspects of Transport Information and Control Systems- Occlusion Method to Assess Visual Demand Due to the Use of In-Vehicle Systems*. Geneva, CH: International Standards Organization.
- Jahn, G., Oehme, A., Krems, J.F., & Gelau, C. (2005). Peripheral detection as a workload measure in driving: Effects of traffic complexity and route guidance system use in a driving study. *Transportation Research Part F*, *8*, 255-275.
- Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data* (DOT HS Rep. 810 594). Washington DC: U.S. National Highway Traffic Safety Administration.
- Kujala, T. (2009). Occlusion technique – Valid metrics for testing visual distraction? In *Proceedings of the European Conference on Cognitive Ergonomics 2009* (pp. 341-348). Otaniemi, FI: VTT Technical Research Centre of Finland.
- Lee, J. D., Regan, M. A., & Young, K. L. (2008). What drives distraction? Distraction as a breakdown of multilevel control. In M.A. Regan, J.D. Lee, & K.L. Young (Eds.), *Driver Distraction: Theory, Effects, and Mitigation* (pp. 41-56). Boca Raton, FL: CRC Press.
- Lenneman, J. K. & Backs, R. W. (2009). Cardiac autonomic control during simulated driving with a concurrent verbal working memory task. *Human Factors*, *51*, 404-418.
- Lleras, A., Rensink, R.A., & Enns, J.T. (2005). Rapid resumption of interrupted visual search: New insights on the interaction between vision and memory. *Psychological Science*, *16*(9), 684-688.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492-527.
- 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* (pp. 2-8). Iowa City, IA: University of Iowa Public Policy Center.
- Moran, T.P. (1981). Guest editor's introduction: An applied psychology of the user. *ACM Computing Surveys (CSUR)*, *13*(1), 1-11.
- Oulasvirta, A. & Saariluoma, P. (2004). Long-term working memory and interrupting messages in human - computer interaction. *Behaviour & Information Technology*, *23*, 53-64.

- Oulasvirta, A. & Saariluoma, P. (2006). Surviving task interruptions: Investigating the implications of long-term working memory theory. *International Journal of Human-Computer Studies*, 64 (10), 941-961.
- Rabbitt, P.M.A. (1984). The control of attention in visual search. In R. Parasuraman and D.R. Davies (Eds.) *Varieties of Attention* (pp. 273-291). Orlando, FL: Academic Press.
- Ratwani, R.M. & Trafton, J.G. (2008). Spatial memory guides task resumption. *Visual Cognition*, 16(8), 1001-1010.
- Ratwani, R.J., Andrews, A.E., McCurry, M., Trafton, J.G., & Peterson, M.S. (2007). Using peripheral processing and spatial memory to facilitate task resumption. *Human Factors and Ergonomics Society Annual Meeting Proceedings*. 51, 244-248.
- Rasmussen, J. (1990). The role of error in organising behaviour. *Ergonomics*, 33, 1185-1199.
- Recarte, M.A. & Nunes, L. (2002). Mental load and loss of control over speed in real driving. Towards a theory of attentional speed control. *Transportation Research Part F*, 5, 111-122.
- Rockwell, T.H. (1988). Spare visual capacity in driving – revisited: New empirical results for an old idea. In M.H. Freeman, C.M. Haslegrave, P. Smith, S.P. Taylor, & A.G. Gale (Eds.), *Vision in Vehicles II* (pp. 317-324). Amsterdam, NED: Elsevier Science Publishers B.V.
- Saariluoma, P., & Oulasvirta, A. (in press). User psychology: Re-assessing the boundaries of a discipline. *Psychology*.
- Schneider, W. & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84 (1), 1-66.
- Shen, Y.J. & Jiang, Y.V. (2006). Interrupted visual searches reveal volatile search memory. *Journal of Experimental Psychology: Human Perception and Performance*, 32(5), 1208-1220.
- Shiffrin, R. M. & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84 (2), 127-190.
- Society of Automotive Engineers (2000). *SAE J2396 Surface vehicle recommended practice, definitions and experimental measures related to the specification of driver visual behavior using video based techniques*. Warrendale, PA: Society of Automotive Engineers.
- Stevens, A., Burnett, G., & Horberry, T. (2010). A reference level for assessing the acceptable visual demand of in-vehicle information systems. *Behavior & Information Technology*, 29, 5, 527-540.
- Underwood, G., & Everatt, J. (1992). The role of eye movements in reading: some limitations of the eye-mind assumption. In E. Chekaluk & K. R. Llewellyn (Eds.), *The Role of Eye Movements in Perception* (pp. 111-169). Amsterdam: North-Holland.
- Underwood, G., Foulsham, T., & Humphrey, K. (2009). Saliency and scan patterns in the inspection of real-world scenes: Eye movements during encoding and recognition. *Visual Cognition*, 17, 812-834.
- Victor, T. W., Engström, J., & Harbluk, J. L. (2008). Distraction assessment methods based on visual behavior and event detection. In M.A. Regan, J.D. Lee, & K.L. Young (Eds.), *Driver Distraction: Theory, Effects, and Mitigation* (pp. 135-165). Boca Raton, FL: CRC Press.

- Victor, T.W., Harbluk, J.L., & Engström, J.A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F*, 8, 167-190.
- Wang, Y., Mehler, B., Reimer, B., Lammers, V., D'Ambrosio, L.A., & Coughlin, J.F. (2010). The validity of driving simulation for assessing differences between in-vehicle informational interfaces: A comparison with field testing. *Ergonomics*, 53, 404-420.
- Wierwille, W. W. (1993). An initial model of visual sampling of in-car displays and controls. In A.G. Gale, I. D. Brown, C. M. Haslegrave, H. W. Kruijse, & S. P. Taylor (Eds.), *Vision in vehicles IV* (pp. 271-279). Amsterdam: Elsevier.
- Wierwille, W.W., Antin, J.F., Dingus, T., & Hulse, M.C. (1988). Visual attentional demand of an in-car navigation display system. In A.G. Gale et al. (Eds.), *Vision in Vehicles II* (pp.307-316), Amsterdam, NL: Noth-Holland Press.
- 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.
- Wikman, A. S. & Summala, H. (2005). Aging and time-sharing in highway driving. *Optometry and Vision Science*, 82, 716-723.
- Zwahlen, H. T., Adams, C. C., & DeBald, D. P. (1988). Safety aspects of CRT touch panel controls in automobiles. In M. H. Freeman, C. M. Haslegrave, P. Smith, S. P. Taylor, & A. G. Gale (Eds.), *Vision in vehicles II* (pp. 335-344). Amsterdam: Elsevier Science Publishers B.V.

Table 1. Point-of-interest search tasks.

Table 2. Visual sampling efficiency, visual demands and driving performance by menu structure (for a device with physical controls) and touch screen scrolling method, means (SEs) of individual values ($n=10$).

Table 3. Significant associations between the lane excursions metrics and the metrics of visual demands and visual sampling efficiency ($N=40$, NLE = number of lane excursions, DLE = duration of lane excursions).

Table 1. Point-of-interest search tasks.

Task #	Task	Path (# of items in the menu)	Menu levels
1	Find the way to the nearest hotel	Options-Search (9)-Hotels	3
2	Find the way to the nearest shop	Options-Search (9)-Shops	3
3	Find the way to the nearest rest area	Options-Search (9)-Automotive (6)-Rest areas	4
4	Find the way to the nearest library	Options-Search (9)-Services (9)-Libraries	4
5	Find the way to the nearest railway station	Options-Search (9)-Transport (4)-Railway stations	4
6	Find the way to a McDonald's restaurant	Options-Search (9)-Restaurants (18)-McDonald's (required scrolling)	4
7	Find the way to a theatre named Kto	Options-Search (9)-Entertainment(2)-Theatres (18)-Kto (required scrolling)	4
8	Find the way to a museum named Dom Jana	Options-Search (9)-Sights (6)-Museums (27)-Dom Jana (required scrolling)	4

Table 2. Visual sampling efficiency, visual demands and driving performance by menu structure (for a device with physical controls) and touch screen scrolling method, means (SEs) of individual values ($n=10$).

	<i>Grid (Exp1)</i>		<i>List (Exp1)</i>		<i>Kinetic (Grid, Exp2)</i>		<i>Buttons (Grid, Exp2)</i>	
<i>Visual sampling efficiency</i>								
Standard deviation of glance durations, s	.66 (.09)		.48 (.03)		.63 (.06)		.52 (.03)	
Maximum glance duration, s	3.70 (.40)*		2.42 (.12)*		3.60 (.36)		2.89 (.10)	
Over-1.6s-glances	25.00 (4.14)*		12.50 (2.30)*		22.80 (3.48)		25.90 (4.10)	
Over-1.6s-glances in curves	8.30 (2.87)*		1.60 (.86)*		2.80 (.59)		3.40 (1.64)	
Over-2.0s-glances	12.40 (2.96)*		5.30 (1.23)*		10.10 (2.16)		10.80 (2.87)	
Over-2.0s-glances in curves	4.30 (1.31)*		.90 (.31)*		.90 (.23)		1.10 (.69)	
Under-0.4-second glances	7.30 (1.44)*		3.40 (.78)*		9.60 (2.13)□		3.40 (.82)□	
Skewness of glance duration distribution	1.09 (.11)**		.66 (.10)**		1.11 (.19)		.75 (.06)	
<i>Visual demands</i>								
Total number of glances	123.50 (11.32)*		93.30 (4.89)*		128.30 (15.14)		142.00 (10.48)	
Average glance duration, s	1.06 (.05)		1.08 (.04)		1.15 (.06)		1.20 (.06)	
Total glance time, s	142.86 (13.39)*		99.60 (4.87)*		144.75 (17.96)		168.12 (10.61)	
<i>Driving performance</i>								
	<i>Base-line</i>	<i>Dual-task</i>	<i>Base-line</i>	<i>Dual-task</i>	<i>Base-line</i>	<i>Dual-task</i>	<i>Base-line</i>	<i>Dual-task</i>
Total number of lane excursions	10.60 (2.47)	18.80 (6.15)	5.40 (1.71)	7.10 (2.01)	8.90 (.88)	16.40 (3.01)	6.30 (1.34)	14.70 (4.04)
Total duration of lane excursions, s	10.73 (3.60)	28.14 (12.18)	4.54 (1.83)	6.59 (2.07)	5.48 (1.40)	18.32 (4.38)	3.99 (1.29)	20.91 (11.24)
Total number of speed maintenance errors	3.40 (.81)	3.90 (1.10)	3.50 (1.24)	2.80 (1.16)	5.70 (1.67)	4.90 (1.44)□	1.90 (.81)	0.80 (.42)□
Total duration of speed maintenance errors, s	20.08 (5.11)	33.01 (11.33)	26.79 (12.66)	34.10 (17.79)	33.62 (10.69)	81.51 (45.12)	9.53 (4.78)	6.17 (3.67)

Note. *: significant difference at .05 level, **: significant difference at .01 level (*Grid* vs. *List, Experiment 1*); □: significant difference at .05 level (*Kinetic* vs. *Buttons, Experiment 2*)

Table 3. Significant associations between the lane excursions metrics and the metrics of visual demands and visual sampling efficiency ($N=40$, NLE = number of lane excursions, DLE = duration of lane excursions).

	<i>Total glance time</i>	<i>Number of glances</i>	<i>Average glance duration</i>	<i>Over-1.6s- glances</i>	<i>Over-2s- glances</i>	<i>Over-1.6s- glances in curves</i>	<i>Over-2s- glances in curves</i>
<i>NLE</i>	.67**	.48**	.33*	.63**	.51**	.88**	.66**
<i>DLE</i>	.61**	.36*	.40*	.69**	.65**	.85**	.72**

Note. *: significant correlation at .05 level, **: significant correlation at .01 level (Pearson's R)