Efficiency of visual time-sharing behavior – The effects of menu structure on POI search tasks while driving

Tuomo Kujala
Agora Human Technology Center
P.O.Box 35
University of Jyväskylä
+358 14 260 4660
tuomo.kujala@jyu.fi

ABSTRACT

In this paper, the effects of two user interface menu structures on a mobile device display, list and grid, are compared in a driving simulation with the measures of visual time-sharing efficiency, visual load, driving performance and secondary task performance. Eighteen participants conducted a set of eight Point-of-Interest (POI) search tasks with the grid- or list-style menus on navigation software during simulated driving. Between-subject analysis revealed that the list-style menu structure supports more efficient and systematic, and thus, safer interaction while driving than the grid-style menu, in terms of time-sharing and total glance time. However, significant effects of the menu structures were not found in secondary task performance, driving performance measured as lane excursions, or in the measures of average duration of, or total number of glances at the display. The results also suggest that the fewer items in a view, the more efficient and safer the interaction in terms of time-sharing. The sensitivity of the time-sharing metrics for revealing tactical level driver distraction in driving simulation can be argued as being at a higher level than the sensitivity of metrics related to lane maintenance, visual load or secondary task performance.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems – *human factors, human information processing.*

General Terms

Measurement, Performance, Design, Reliability, Experimentation, Human Factors, Theory.

Keywords

Driver distraction, time-sharing, visual interaction, displays, menu structures, workload, visual load, driving performance, levels of control, tactics, strategies.

1. INTRODUCTION

The safety effects of in-vehicle information system (IVIS) use while driving is a topic that is gaining more and more attention these days because of the fast development of mobile technology and services [12]. In this line of research, driver distraction is the key concept defined by Lee, Young, and Regan [10] as *a diversion of attention away from activities critical for safe driving toward a competing activity.*

The experimental approach on studying driver distraction has been an area of interest in human factors research since the 1980s. Driving simulation studies have been frequently used in order to avoid real crash risk (see e.g. [3]). A popular paradigm in this line of research has been based on the measurement of driver workload and driving performance at the level of operational control of the vehicle [9]. The basic problems with interpreting the results of these experiments often reside in the not-self paced and time-pressured tasks, and subsequently in the absence of participants' possibilities to prioritize the driving over secondary tasks. The external validity of the conclusions can often be questioned (e.g., [7][17], see also [6]). These studies are valuable for revealing capacity limitations of the drivers in a dual-task situation. However, they do not necessarily tell us if the drivers are able to overcome their capacity limits with tactical behaviors in real traffic to maintain a sufficient level of driving performance.

Recently, new perspectives and models for studying driver distraction on multiple levels have been proposed [9][14]. Lee, Regan, and Young [9] introduced the model of driver distraction comprising of breakdowns at the operational, tactical and strategic levels of control in dual-tasking while driving based on Michon's [11] three-level model of driving behavior. This model induces new types of challenges for experimental research; how can breakdown in control be measured on the levels of tactical and strategic control? These are not necessarily in direct relation to task workload or to the lapses of vehicle control at the level of operational control [9].

In this paper, while focusing on interaction with visual IVIS displays, we can ask; what kinds of display design solutions could support drivers' tactical and strategic skills in overcoming their visual capacity limitations? Task predictability, interruptability, resumability, and ignorability have all been acknowledged as important secondary task qualities for promoting traffic safety [1] [5][9][16], but guidelines, such as the European Statement of Principles [1], are not specific in defining how particular display properties relate to these aspects and how to measure them. Awareness of task demands and one's own capabilities, i.e. situation awareness, are related concepts that are highly relevant for drivers' tactical and strategic abilities in a dual-task situation. How can these be measured in an objective way?

AutomotiveUI'09, September 21-22, 2009, Essen, Germany
ACM 978-1-60558-571-0/09/009

The typical measures in dual-tasking experiments focus on measuring driver workload (e.g. visual load) or performance at the operational level of control (e.g. driving performance). Traditional measures of visual load focus on the average glance durations or on the total glance durations (i.e. total glance time, tgt) and total number of glances at the display. However, it has been observed that in general, drivers tend to keep the average glance durations below 1.6 seconds in all circumstances and increase the frequency of glances instead of increasing the lengths of individual glances while the visual secondary task demands increase [20]. This is natural behavior, if we acknowledge that drivers, *in general*, try to behave as rational and intentional human beings in traffic.

In this paper, visual time-sharing, or time-sharing in short, is defined as allocation of visual attention in time between tasks. Time-sharing-metrics have been suggested and also utilized to a minor degree to provide information on the glance duration distributions towards a visual secondary task, and thus, on the total efficiency of the allocation of visual attention [2][4][18][21][22]. Very short glances at an in-vehicle display can indicate inefficient search behaviors, as well as rare but significantly long glances that can also increase the level of crash risk [4]. Thus, time-sharing metrics could presumably provide us information on driver distraction at the tactical level of dual-task control. For example, a significant difference in the variance of glance durations on two display designs could tell us that the design with lower variance gives better support for controlled visual search behavior, given the same variability of the driving task's visual demands. In addition, the significance of even one "overlong" glance at an in-vehicle display in the wrong situation cannot be emphasized enough [4]. The traditional measures of average or total glance durations cannot provide us with information on the frequencies of these often rare occasions.

For industrial purposes, fast but sensitive and reliable methods for revealing differences in the distraction potentials of visual IVIS displays are obviously required. Sensitivity means that the metrics can discriminate between designs reliably with statistical significance already with small sample sizes, and thus enables cost-efficient studies. The methods should also provide us with information about driver behavior on multiple levels of driver distraction [9], not merely on the level of operational control, for enabling higher external validity of the conclusions.

The experiment presented in this paper relates to a real design problem in the design of navigation software for a mobile device. The problem goes; which menu structure should be used in the driving mode of the software: list or grid (see Figure 1)? Does this decision have some potential effects on traffic safety? Intuitively, one could argue that the grid-style menu supports faster interaction by shorter paths to more items than the list-style menu. In addition, larger icons can be used and a single view can show more items at once than the list-style menu, thus enabling lower menu structures. All these aspects could support faster, and thus, perhaps safer interaction while driving. On the other hand; the list-style menu could support more predictable interactions because of the more straight-forward two-way movements in the menu. However, in bench-tests without driving, the interaction with either menu does not seem to be significantly more complex than with the other.

Figure 1. Two alternative menu structures, list and grid, for the driving mode of the navigation software.

In this paper, the following questions are addressed:

-Which menu structure, grid or list, supports safer interaction with a mobile device while driving? Is there a significant difference between the two designs with any of the measures?

-Do the amount of items in a view, or the levels of menu, have moderating effects on the previous issues?

In addition, the sensitivity of lane maintenance, secondary task performance, visual load, and time-sharing metrics are compared. What types of measures could indicate significant effects already with small sample sizes? Are the metrics of time-sharing efficiency suitable and sensitive enough for assessing distraction effects of in-vehicle display designs at the level of tactical and strategic control?

2. METHOD

Two hypotheses were made prior the experiment based on our earlier research. Firstly, interaction with the list-style menu is assumed to be safer while driving than with the grid-style menu, because it could support more systematic visual interaction. The visual demands of the interaction are thus more easily learnable, and thus more predictable, interruptible and resumable than when interacting with the grid-style menu. This should be visible with the measures of time-sharing efficiency, but not necessarily with the measures of lane maintenance, visual load or secondary task performance. We expected larger variances in glance duration distributions, larger maximum glance durations, and greater amounts of very long, as well as very short glances towards the display with the grid-style menu. Secondly, the fewer items in a view and the lower the menu structure, the more efficient the interaction is supposed to be in terms of time-sharing.

2.1 Participants

Volunteers were invited to participate through public university e-mail lists. The 6 female and 12 male right-handed participants were from the ages of 20 to 35 years old, and had normal or corrected vision. All had a valid driving license and possessed lifetime driving experience of at least 10 thousand kilometers, ranging from 10,000 to 500,000 km. Drivers with a very low level of experience and aged drivers were not selected to the sample for mitigating the known effects of low level of driving experience [21] and aging [22] on time-sharing efficiency. The experiment was conducted in Finnish with fluent Finnish-speakers.

Participants were randomly selected from the volunteers, but they were divided in two pair-matched groups according to gender, levels of driving experience and age (see Table 1). The group with the grid-style menu had an average lifetime driving experience of 103,000 km (SD=159), and an average age of 25.1 years (SD=2.8). For the *List*-group the corresponding averages were 95,000 km (SD=124) and 25.7 years (SD=5.0).

Table 1. Classes of pair-matched participants

Number of participants	Driving experience (thousand km)	Age
6	<20	20-25
2	<20	26-35
2	20-50	20-25
2	50-100	20-25
2	>100	20-25
4	>100	26-35

2.2 Tools and environment

The experiment was conducted in the three-display driving simulation environment of the Agora User Psychology Laboratory (see Figure 2).

Figure 2. The driving scene from a participant's point of view.

The central equipment included consent forms, Nokia N95 8GB mobile device with 2.8" display in a dashboard holder, SMI iView X HED helmet-mounted eye-tracking system with 50Hz sampling rate, two video cameras for recording the driving scene with sound and for back-upping the eye-tracking, as well as two laptops for capturing the video material. The distance between participants' eyes and the windscreen projected driving scene was fixed at ca 100 cm, but the distance of the pedals and the steering wheel with the device holder from the participant were adjustable. Thus, the mobile device's distance from the participant's eyes varied between 55 to 70 centimeters depending on arm lengths.

The driving simulation software is an open-source based car simulation of which motion formulae is based on actual engineering documents from the Society of Automobile Engineers (www.racer.nl). The trials were driven with a simulated Ford Focus with automatic shifting on a road-like environment simulating the Polish countryside. A simulated racetrack was used for practice. The driving scene was projected onto the wind screen of the fixed-base vehicle cockpit and included a speedometer and a tachometer.

2.3 Design and procedure

The experimental design was a mixed-factorial design (see Table 2). The menu structure was a between-subject variable and the levels of menu, and the number of items in the view, were the within-subject variables.

Table 2. The experimental design

Menu (between-subject)	Levels (within-subject)	Items (within-subject)
List	3	2
Grid	4	4
	>4	6
		9

The experiment started with the signing of a consent form, and by receiving general instructions. Practice in driving on a looped track of around 5 minutes was provided for the participant. After the rehearsal the participant completed a baseline driving task of around 10 minutes for getting more practice and for baseline-dual-task driving performance comparisons. The participant got to complete one search task without driving with the search tasks on the grid- or list-style menus before the dual-task trial. The dual-task trial lasted for 6 to 10 minutes depending on the participant's task completion times. After driving, the participant was interviewed in order to explore the participants' strategy space and to classify the drivers' ways of interacting. Both menu structures were shown to the participants during the interviews. The main questions of interest in the interviews were: "Did you feel time-pressure or need to hurry in the search tasks?"; "How did you perform the search tasks; did you have or did you develop certain ways of interacting during the trial?"; "Which menu structure would you prefer to use while driving?"; and finally, "Could you imagine yourself conducting this type of search activity while driving?".

The driving task instructions were to keep the velocity of the vehicle between 40-60 km/h, and to keep the two Head-Up-Display meters between the white lane markings. The participant was also instructed to stop the vehicle immediately if he/she saw a deer. Driving practice included a deer, but the actual trials did not. However, the participants were not made aware of this beforehand. There was oncoming traffic in the form of four cars at preset points on the road.

The search tasks were *self-paced* and the participant was instructed to keep *priority on driving*. Driving task priority was emphasized by promising 10 movie tickets in total to the most accurate drivers. Driving task accuracy was defined as the total duration spent out of the lane or above/below the instructed speed zone. Tasks were given verbally by the experimenter while driving, allowing for a very short pause of a few seconds between tasks after a successful task. The participant could ask the task to be repeated with saying “repeat”, if he/she forgot or did not hear the task.

Participants were given the scenario that they are travelling in the Polish countryside by car and searching for Points-of-Interest (POIs) nearby. The search tasks are listed in Table 3. The number of items in a view varied within the tasks depending on the level of the menu. Task orders were randomized.

Table 3. The search tasks

Task #	Task (path (# of items))	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 the nearest McDonald’s (Options-Search(9)-Restaurants(18)-McDonald’s [required scrolling])	>4
7	Find the way to the theatre named Kto (Options-Search(9)-Entertainment(2)-Theatres(18)-Kto [required scrolling])	>4
8	Find the way to the museum named Dom Jana (Options-Search(9)-Sights(6)-Museums(27)-Dom Jana [required scrolling])	>4

2.4 Variables and analysis

Independent variables included the menu structure, the levels of menu, and the number of items. Efficiency of time-sharing, visual load, driving performance and search task performance were selected as dependent measures.

Time-sharing efficiency was measured by the maximum and standard deviations of glance durations (at the display), by the frequency of over-1.6-second and over-2-second glances in total and in curves, and by the frequency of under-0.4-second glances. 1.6 seconds has been observed to be the limit under which drivers generally prefer to keep their glances at in-vehicle displays in all circumstances [20]. Over-2-second long glances have been observed to increase crash risk and frequency of near crash situations in real traffic [8]. Additionally, we wanted to include a measure of situation awareness. The metrics of overlong glances while driving in curves served this purpose. The movement of gaze from the driving scene to the display and back was scored into the glance duration, and as such, under 0.4 second glances leave very short time for gathering any useful information from the display, especially if assuming some task set switch costs (see e.g. [13]). A typical shift of gaze between the display and the driving scene took 160 ms. The effects of the levels of menu and the number of items were analyzed for maximum and standard deviations of glance durations when applicable (enough glances). Interaction effects of menu structure and the within-subject-variables on these measures were also analyzed.

Visual load was measured by the total number and mean duration as well as total duration of glances at the display. Driving performance was measured as total frequency and duration of lane excursions, and additionally involved baseline-dual-task comparisons. The within-subject effects of the levels of menu on driving performance were excluded in the analysis. Total frequency and duration of speed area violations were scored automatically from the simulation log file for the accuracy comparisons between participants. Secondary task performance was measured as frequency of errors and task completion times with driving excluded, that is, total glance times at the display by task. Error in a search task was defined as a wrong selection.

Effort was invested to control some undesired variables. There was an effort to accommodate for learning effects and individual differences in skills via driving practice and practice for the search task. As mentioned, the menu-groups were balanced by gender, driving experience and age. Order effects were eliminated by randomizing orders of the search tasks (5 different orders, same orders for the pairs). Driving task difficulty while dual-tasking was controlled by random task starting points, which depended on the participants’ performance. In addition, every other participant in the group drove the same road in the dual-task trial as the others, but in the opposite direction. This kept the driving task demands (road curvature) at the same level for everyone but gave more randomness to the task starting points. The driving speed was kept fixed between 40-60 km/h by instructions. Movie tickets were promised to the most accurate drivers in order to make the participants prioritize the driving and to encourage greater effort, giving the absence of real danger in a driving simulation. The deer observation task was instructed to make the participants observe the environment in a more natural way than merely observing the lane markings and the speedometer.

Noldus Observer XT software was used for scoring behaviors frame-by-frame (25 frames per second), for search task performance, lane excursions, and eye-movements. A glance at the secondary task display was scored following the SAE J2396 definition [15]. The analysis of overlong glances in curves was done via an automatic script that compared the steering wheel movements recorded in the log file of the driving simulation to the synchronized eye-tracking data file. The limit for driving on a curve was defined to be the absolute value of 1.00 or more of the steering wheel position in terms of the simulation’s log file data, in which 0.00 was the calibrated center point. The frequency and durations of lane excursions were analyzed for equal journey lengths between the two trials. Due to there being an end in the road, there was a time limit of about 10 minutes for the completion of the search tasks. Three participants did not have enough time to start the last tasks in their trials. This was taken into account in the analysis by excluding the corresponding task data from their pairs in the other group. The time limit was not instructed to the participants. For statistical analysis, two-tailed *t*-tests were used for between-subject comparisons and mixed-model ANOVA (menu x level, 2 x 3; menu x items, 2 x 4) to find within-subject as well as interaction effects. An alpha level of .05 was used in the statistical testing. The interviews were analyzed from the videos and the participants’ answers were classified.

3. RESULTS

3.1 Driving performance

The means for the total number of lane excursions were 7.78 (SEM=2.12) for the list and 20.11 (SEM=6.71) for the grid. Correspondingly, the means for the total duration of lane excursions were 7.26 s (SEM=2.19) for the list and 30.26 s (SEM=13.41) for the grid. However, these differences were not statistically significant (total number: $t(16)=1.75$, $p=.110$; total duration: $t(16)=1.69$, $p=.130$).

Despite of the higher level of practice in driving after the baseline driving trial, there was a significant effect of the dual-task condition on the total number of lane excursions ($F(1,16)=4.92$, $p<.050$, see Figure 3). Analysis of speed variations did not reveal high numbers of significant speed zone excursions and speed was not included in the analyses.

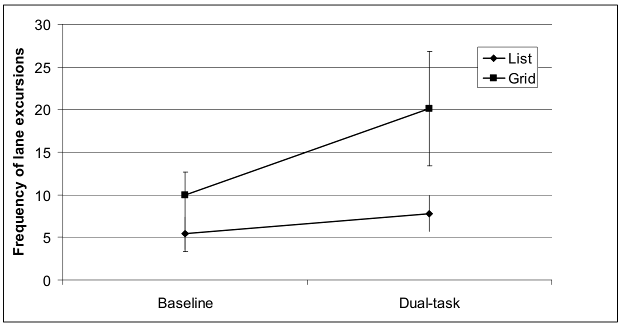

Figure 3. Total number of lane excursions by trial (N=18). Means and Standard Errors of Means.

3.2 Visual load

Total glance times was the only measure of visual load that showed significantly larger values for the grid-style menu ($t(16)=2.91$, $p<.050$, see Table 4). The mean glance lengths at the displays were similar and indicate safe *average* visual behavior [20].

Table 4. Visual load (N=18), means (SEMs)

Menu	Total glance time, s	Total number of glances	Average glance duration, s
List	98.20 (5.22)	93.56 (5.46)	1.07 (.05)
Grid	144.10 (14.90)	121.33 (12.43)	1.06 (.04)

3.3 Time-sharing efficiency

The effects of the menu structures on the participants' time-sharing efficiency are illustrated in the Figure 4.

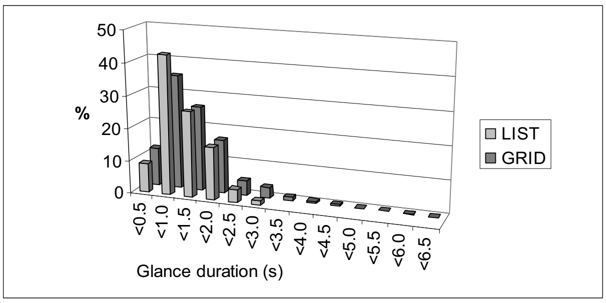

Figure 4. Glance duration distributions (N=18).

The maximum glance durations at the grid-style menu were significantly larger than at the list-style menu ($t(16)=2.93$, $p<.050$, see Figure 5). The means for the individual standard deviations of glance durations were also larger for the grid-style menu, but the difference was not statistically significant ($t(16)=1.93$, $p=.072$) in these sample sizes.

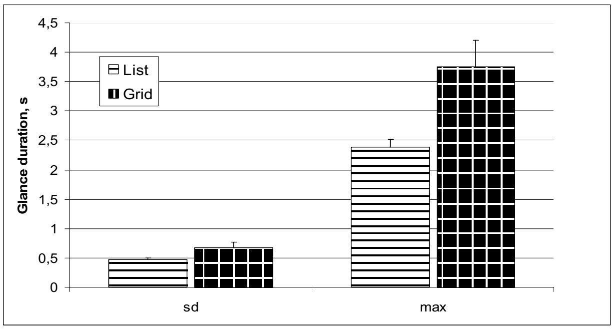

Figure 5. Standard deviation of and maximum glance durations (at the display, N=18). Means and SEMs.

There was a significant effect of the menu structure on the frequencies of over-1.6s-glances ($t(16)=2.82$, $p<.050$) and over-2.0s-glances ($t(16)=2.12$, $p<.050$, see Figure 6). The larger frequencies of over-1.6s-glances ($t(16)=2.22$, $p<.050$) and over-2.0s-glances ($t(16)=2.41$, $p<.050$) in curves indicates, that the participants using the grid-style menu also did significantly more of these glances while they were not driving on a straight road. In addition, there was a significantly larger number of under 0.4-second-glances on the grid-style menu ($t(16)=2.12$; $p<.050$). A summary of the time-sharing metrics is presented in the Table 5.

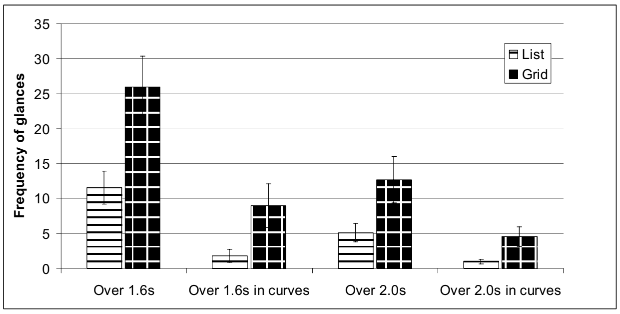

Figure 6. Frequencies of overlong glances in total and while driving in a curve (N=18). Means and SEMs.

Table 5. Time-sharing metrics (N=18), means (SEMs)

Measure	List	Grid
Standard deviation of glance durations, s	.47 (.03)	.67 (.10)
Maximum glance duration, s	2.38 (.13)	3.75 (.45)
Over-1.6s-glances	11.56 (7.02)	25.89 (4.52)
Over-1.6s-glances in curves	1.78 (.81)	9.00 (3.11)
Over-2.0s-glances	5.11 (1.36)	12.67 (3.30)
Over-2.0s-glances in	1.00 (.33)	4.56 (1.44)

curves		
Under-0.4s-glances	3.22 (.85)	7.00 (1.57)

3.3.1 *Number of items*

The number of items on a view had a significant effect on the maximum glance durations at the display (F(3,14)=26.02, p=.000, see Figure 7). Between-subject effects of the menu structure were not significant, but there was significant interaction effects between the menu structure and the number of items (F(3,14)=3.70, p<.050). There were significantly larger maximum glance durations on 9 item views with the grid-style menu compared to the list-style menu, *t*(16)=2.93, p<.050.

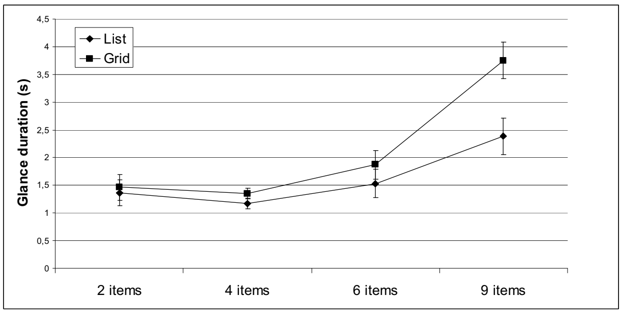

Figure 7. Maximum glance durations by the number of items (N=18). Means and SEMs.

On the task level, the standard deviations of glance durations in task 8 (27 items in the last menu) were significantly larger with the grid-style menu (M=.74, SEM=.12) than with the list-style menu (M=.48, SEM=.04), *t*(14)=2.14, p<.050.

3.3.2 *Number of menu levels*

The number of menu levels had a significant effect on the maximum glance durations at the display (F(2,15)=8.67, p<.010, see Figure 8).

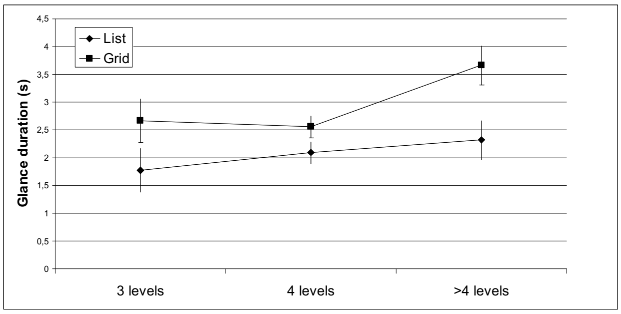

Figure 8. Maximum glance durations by the number of menu levels (N=18). Means and SEMs.

There was also a significant between-subject effect of the menu structure on the maximum glance durations (F(1,16)=5.32, p<.050). However, significant interaction effects of the variables were not found. Standard deviations of glance durations did not show significant effects of the menu levels, but indicated significantly larger values for the grid-style menu (M=.72, SEM=.10) than for the list-style menu (M=.47, SEM=.03) in the >4 level tasks, *t*(16)=2.38, p<.050.

3.4 Search task performance

The menu structure did not significantly affect task performance on any other task than #2 (tgt: *t*(14)=2.28, p<.050). The most difficult tasks in total seemed to be the tasks 7 and 8 (see Table 6).

Table 6. Task times and errors in the search tasks

Task #	N	Menu	Task time (as tgt, mean (SEM), s)	Errors (sum)
1	9	List	7.86 (1.40)	0
1	9	Grid	9.65 (.78)	2
2	8	List	6.81 (1.28)	1
2	8	Grid	16.94 (4.26)	6
3	9	List	13.01 (2.56)	2
3	9	Grid	18.31 (3.56)	10
4	9	List	12.80 (1.92)	3
4	9	Grid	13.65 (1.46)	4
5	9	List	7.48 (.73)	0
5	9	Grid	8.66 (1.11)	1
6	8	List	13.71 (.91)	0
6	8	Grid	26.91 (5.95)	4
7	9	List	19.62 (2.35)	9
7	9	Grid	18.77 (4.42)	4
8	8	List	21.58 (2.89)	8
8	8	Grid	40.59 (10.66)	23

3.5 Interviews

The participants reported that the development of “sense of touch” and memory helped them in doing some of the movements without looking at the device, e.g. in the event of the repeated ‘Search’-selection. This observation could be confirmed from the video material. Some participants jumped directly between the first and the last item in the menu; in these cases the participants reported that the list better supported the awareness of location in the menu (notice: there was no scroll bar in the scrollable submenus). All of them reported they mainly read the texts. Icons were reported to have been helpful only in the identification of certain targets, e.g. restaurants. Some felt that the texts were easier to distinguish in the grid. There were notes of difficult menu hierarchies in some tasks, especially in deciding to which submenu rest areas, theaters, and museums belonged to.

Twelve out of 18 participants reported that they did not feel time-pressure while completing the search tasks. Four participants reported some time-pressure, but they felt it was self-induced. Two participants reported that they felt time-pressure while completing the tasks, but that they also tried to prioritize driving. When asked for preference, 10 participants preferred the list, 6 preferred the grid, one preferred a combination of grid and list depending on the view, and one felt it did not make a difference. Thus, all of them were not fully aware of the risks of the grid-style menu. Finally, 15 out of the 18 participants could imagine themselves conducting these types of search tasks while driving.

4. DISCUSSION

In this paper, we studied the sensitivity of the metrics of lane maintenance, secondary task performance, visual load, and time-sharing efficiency, to reveal lapses in participants' tactical visual behaviors while driving and dual-tasking with visual secondary tasks. This was achieved through a case study in which we compared the effects of mobile software's menu structures on these measures.

As hypothesized, the collected time-sharing data suggests that the unsystematic nature of visual search and movement that the grid-style menu offers may lead to less efficient, less predictable, less resumable and less interruptible interactions while driving than the interaction with the list-style menu. This was observed especially with displays featuring over 6 items. Although the lane maintenance metrics provided some hints of this, they do not seem to be very sensitive to revealing risky visual behavior or differences in the distraction potential of visual display properties at this level. A possible reason for the non-significant effects of menu structures on lane maintenance could be that the drivers' performance at the level of operational control of the vehicle does not need to be in direct relation with the lapses of tactical level behavior [9]. Driving performance is, besides visual behavior, highly dependent on other simultaneous factors, such as the curvature of the road at any one time (see [4]).

Nor does the metrics of visual load seem to be highly sensitive in revealing occasional risky visual behavior. It was only the total glance times that indicated significantly faster performance for the list-style menu. However, neither does this measure of visual load have to be in a direct relationship to risky visual behavior in all cases. It is highly possible that the total glance times for two secondary tasks are the same, but the glance duration distributions are very different. The crash risk potential in this case is not about the visual load, i.e. the total or mean amount of visual attention required for a secondary task, but about how efficiently you can allocate your visual attention in time between driving and the visual secondary task (see [4]). This time-sharing behavior is a critical component of successful driving, and not merely for lane keeping and collision avoidance, but also for detecting and preparing for potential, unexpected threats in time.

Analysis of search task performance revealed that the difficulty (i.e. complexity) of the secondary task is not necessarily the main factor for unsafe time-sharing behavior. In other words, task complexity can be at the same level between two visual secondary tasks, but still they can have different effects on time-sharing efficiency, and consequently, on potential crash risk [4].

Some design recommendations can be inferred from the results. The list-style menu seems to support safer interaction while driving than the grid-style menu. However, it should be noted, that this recommendation at this point only applies to mobile devices with no touch screen. The selection of an item on a touch screen display can be much more straight-forward. The grid-style menu's higher potential for very short and inefficient, as well as overlong glances in relation to the driving situation, can be explained by the more unsystematic steps of interaction. Thus, IVIS display designs should support systematic ways of interaction. Furthermore, the results suggest that the maximum of between 6 to 9 items on a single view can make the information search less risky. Thus, prioritization of what kind of POIs are the most often needed by the drivers, and implementation of only these, could be beneficial. On the other hand; the data seems to suggest, that the lower the menu structure, the better, and these two requirements are often in contradiction with each other in versatile software. However, the effects of the number of menu levels should be examined more closely in different experimental design, because in this experiment the observed effects of menu levels can be explained by the large number of 9 item views in the over 4 level tasks. An interesting finding was that in a short time the participants learned to find and select the often repeated functions of the software (Options and Search) without any visual attention. One possible reason for the better time-sharing efficiency with the list-style menu could relate to this finding in that the participants were able, after locating the target item, to quickly calculate the required steps to the item, and perform the movement without looking at the device. This was much more difficult for the grid-style menu. Moreover, these findings suggest that the sense of touch is worth supporting whenever possible. The finding relates to Wickens' [19] theory of multiple resources, suggesting that tasks which occupy different sense modalities can often be efficiently performed simultaneously. The participants seemed to be aware of this and utilized this information tactically.

The current findings have some methodological significance. The time-sharing-metrics utilized here seem to be suitable for measuring task predictability, resumability and interruptability. They also seem to be sensitive enough for discriminating between safer and not-as-safe, as well as between efficient and less efficient IVIS visual display designs, reliably already with small sample sizes. The efficiency of time-sharing seems to be more closely related to the tactical abilities of drivers than to the complexity or visual load (i.e. workload) of the secondary task. In this sense, we can discuss of a kind of "paradigm shift" in driver distraction research with experimental techniques. In this line of research, instead of asking; does the secondary task overload the driver, we ask; what qualitative features of IVIS display designs can support drivers' tactical and strategic abilities? The target of analysis in experimentation at this level resides in the efficiency of drivers' visual time-sharing behavior, not in the visual load or driver's performance at the operational level of vehicle control [9]. For example, glance duration distributions, or in particular, the frequencies of very short glance durations are rarely looked at in contemporary distraction research [12].

The experiment invoked new questions of which some are already under on-going research. By utilizing the time-sharing metrics, the comparison of touch screen vs. non-touch screen devices with the grid-style menu could reveal whether the main source of the observed distraction is related to moving in the menu, or merely to searching for information in the grid-style views. Another interesting issue under current experimentation is the means of scrolling on a touch screen with menus consisting of more items than the display can hold at a time. Overall, a large amount of similar issues exists that should be tested with the time-sharing metrics. The cumulating database could also serve for comparisons between visual IVIS designs. Moreover, the metrics themselves may be further developed, and validated e.g. with field studies. Future research efforts should especially include further development of the metrics for situation awareness and systematicity of drivers' visual search behavior on IVIS displays.

Finally, please notice that we do not take any account with these results as to whether this type of activity while driving is risk-free or not, even after a large amount of practice. The focus of inquiry was on first-touch experience with the software while driving. We wanted to see how the different menu structures affect visual search behavior with unfamiliar contents, because search for POIs involves typically unseen, dynamic contents depending on the driver's location. However, the driving speed was relatively low in the experiments, *40-60 km/h* and all of the participants were fairly *experienced young drivers*. At higher speeds, or with less experienced [21] or aged drivers [22], the safety risks of the secondary tasks will presumably be greater than observed. With these experiments, we wanted only to reveal which menu interaction style is *safer* while driving – the list or the grid-style – and get support for our hypotheses as to why it is so. We recommend that drivers are made aware of the potential risks of system use while driving. We further encourage them to be fully stationary while searching for locations, whenever possible.

5. ACKNOWLEDGMENTS

This research was sponsored by Theseus, a research project on human-technology interaction design funded by TEKES (Finnish Funding Agency for Technology and Innovation). The driving simulation environment was developed with a grant from the Henry Ford Foundation. The author would like to thank Professor Pertti Saariluoma and research assistant Ari-Matti Saari for their help in planning and conducting the experiments. Credits for the editing of spelling go to Rebekah Rousi. The author wants to express gratitude especially to Mikko Nirhamo and Timo Tokkonen at Nokia for the valuable collaboration in making the experiments possible.

6. REFERENCES

[1] Commission of the European Communities 2006. Recommendation on safe and efficient in-vehicle information and communication systems: Update of the European statement of principles on human machine interface (Report C(2006) 7125 final). Brussels, BEL: Commission of the European Communities.

[2] Hoffman, J. D., Lee, J. D., McGehee, D. V., Macias, M., and Gellatly, A. W. 2005. Visual sampling of in-vehicle text messages: Effects of number of lines, page presentation, and message control. Transport. Res. Rec. 1937, 22-30.

[3] Horrey, W. J. and Wickens, C. D. 2006. Examining the impact of cell phone conversations on driving using meta-analytic techniques. Hum.Factors 48, 196-205.

[4] Horrey, W. J. and Wickens, C. D. 2007. In-vehicle glance duration: Distributions, tails, and model of crash risk. Transport. Res. Rec. 2018, 22-28.

[5] 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.

[6] Jamson, A. H., Westerman, S. J., Hockey, G. R. J., and Carsten, O. M. J. 2004. Speech-based e-mail and driver behavior: Effects of an in-vehicle message system interface. Hum. Factors 46, 625-639.

[7] Just, M. A., Keller, T. A., and Cynkar, J. 2008. A decrease in brain activation associated with driving when listening to someone speak. Brain Res. 1205, 70-80.

[8] Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., and 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.

[9] Lee, J. D., Regan, M. A., and Young, K. L. 2008. What drives distraction? Distraction as a breakdown of multilevel control. In Driver Distraction: Theory, Effects, and Mitigation. M.A. Regan, J.D. Lee, and K.L. Young, Eds. CRC Press, Boca Raton, FL, 41-56.

[10] Lee, J. D., Young, K. L., and Regan, M. A. 2008. Defining driver distraction. In Driver Distraction: Theory, Effects, and Mitigation. M.A. Regan, J.D. Lee, and K.L. Young, Eds. CRC Press, Boca Raton, FL, 31-40.

[11] Michon, J. A. 1985. A critical view of driver behavior models: What do we know, what should we do? In Human Behavior and Traffic Safety. L. Evans and R. C. Schwing, Eds. Plenum Press, New York, NY, 485-520.

[12] Regan, M.A., Lee, J.D., and Young, K.L. 2008. Driver Distraction: Theory, Effects, and Mitigation. CRC Press, Boca Raton, FL.

[13] Sakai, K. 2008. Task set and prefrontal cortex. Annu. Rev. Neurosci. 31, 219-245.

[14] Sheridan, T. B. 2004. Driver distraction from a control theoretic perspective. Hum. Factors 46, 587-599.

[15] 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.

[16] Society of Automotive Engineers 2004. Navigation and route guidance function accessibility while driving, SAE recommended practice J2364. Warrendale, PA: Society of Automotive Engineers.

[17] Sodnik, J., Dicke, C., Tomazic, S., and Billinghurst, M. 2008. A user study of auditory versus visual interfaces for use while driving. Int. J. Hum.-Comput. St. 66, 318-332.

[18] Victor, T. W., Harbluk, J. L., and Engström, J. A. 2005. Sensitivity of eye-movement measures to in-vehicle task difficulty. Transport. Res. F-Traf. 8, 167-190.

[19] Wickens, C.D. 2002. Multiple resources and performance prediction. Theoretical Issues in Ergonomics Science 3 (2), 159-177.

[20] Wierwille, W. W. 1993. An initial model of visual sampling of in-car displays and controls. In Vision in Vehicles IV, A.G. Gale, I. D. Brown, C. M. Haslegrave, H. W. Kruysse, and S. P. Taylor, Eds. Elsevier Science Publishers B.V., Amsterdam, NL, 271-279.

[21] Wikman, A.S., Nieminen, T. and Summala, H. 1998. Driving experience and time-sharing during in-car tasks on roads of different width. Ergonomics 41 (3), 358-372.

[22] Wikman, A. S. and Summala, H. 2005. Aging and time-sharing in highway driving. Optometry Vision Sci. 82, 716-723.