

# Driving while using a smartphone-based mobility application: Evaluating the impact of three multi-choice user interfaces on visual-manual distraction



N. Louveton <sup>a,\*</sup>, R. McCall <sup>a,c</sup>, V. Koenig <sup>a,b</sup>, T. Avanesov <sup>a</sup>, T. Engel <sup>a</sup>

<sup>a</sup> Interdisciplinary Centre for Security, Reliability and Trust, Weicker Building, Université du Luxembourg, 4, Rue Alphonse Weicker, L-2721 Luxembourg

<sup>b</sup> ECCS Research Unit, Campus Walferdange, Université du Luxembourg, Maison des Sciences Humaines, 11, Porte des Sciences, L-4366 Esch-sur-Alzette, Luxembourg

<sup>c</sup> ERIN Department, Luxembourg Institute of Science and Technology, 5, avenue des Hauts-Fourneaux, L-4362 Esch-sur-Alzette, Luxembourg

## ARTICLE INFO

### Article history:

Received 31 October 2014

Received in revised form

23 November 2015

Accepted 28 November 2015

Available online xxx

### Keywords:

Driving simulator

Multi-touch application

Visual-manual distraction

## ABSTRACT

Innovative in-car applications provided on smartphones can deliver real-time alternative mobility choices and subsequently generate visual-manual demand. Prior studies have found that multi-touch gestures such as kinetic scrolling are problematic in this respect. In this study we evaluate three prototype tasks which can be found in common mobile interaction use-cases. In a repeated-measures design, 29 participants interacted with the prototypes in a car-following task within a driving simulator environment. Task completion, driving performance and eye gaze have been analysed. We found that the slider widget used in the filtering task was too demanding and led to poor performance, while kinetic scrolling generated a comparable amount of visual distraction despite it requiring a lower degree of finger pointing accuracy. We discuss how to improve continuous list browsing in a dual-task context.

© 2016 Elsevier Ltd and The Ergonomics Society. All rights reserved.

## 1. Introduction

Advances in the area of digital mobility technologies (Fishman, 2012) have led to an increasing number of displays and interactive systems within cars. Although the safety risk of using mobile devices while driving is well documented (Alm and Nilsson, 1995; Young et al., 2007; Treffner and Barrett, 2004; Horberry et al., 2006). Innovative applications targeting In-Vehicle Infotainment Systems (IVIS) are likely to place an increasingly high workload on the driver. Despite the trend towards the integration of IVIS specific devices into high-end car models, smartphones remain a natural and mainstream method of delivering new mobility and trip-centric applications. An example of these new mobility-centric services are the applications being developed within the i-Gear project (McCall and Koenig, 2012; McCall et al., 2013). This project intends to provide drivers with real-time alternative mobility choices in order to avoid traffic jams. For instance, two use-cases

are envisioned: the possibility to share a ride with a friend or to engage in an alternative activity, which will consume a certain amount of time but will steer the driver away from the peak hours or the congested roads. The ultimate goal of this application is to improve the traffic situation in cities with a high number of commuters.

One major challenge is that providing the driver with more choices displayed on a smartphone will increase visual-manual distraction; as it requires them to interact with visual content and to provide manual inputs. Such an increase in visual-manual distraction while driving is very likely to reduce driving performance and safety (Green, 2004). Visual-manual interactions on modern mobile devices are generally performed through multi-touch gesture inputs on a graphical display (i.e., any finger gestures used on a hand-held or IVIS-specific device); this induces visual-manual distraction and needs to be explored more thoroughly. While general behavioural laws (Fitts, 1954; Hick, 1952) can provide guidance for interface design, specific types of gesture interactions or visual presentations may impact upon the driver's performance and this needs to be empirically assessed.

Different methods for interacting with list-based applications have been assessed. For instance, Kujala (Kujala, 2009) studied the

\* Corresponding author.

E-mail addresses: [nicolas.louveton@uni.lu](mailto:nicolas.louveton@uni.lu) (N. Louveton), [roderick.mccall@list.lu](mailto:roderick.mccall@list.lu) (R. McCall), [vincent.koenig@uni.lu](mailto:vincent.koenig@uni.lu) (V. Koenig), [tigran.avanesov@uni.lu](mailto:tigran.avanesov@uni.lu) (T. Avanesov), [thomas.engel@uni.lu](mailto:thomas.engel@uni.lu) (T. Engel).

effect of grid and linear presentations of icon lists on driving performance and gaze behaviour. They found that the linear list layout results in better visual search patterns and as a consequence should be safer. Scrolling mode was also assessed by (Lasch and Kujala, 2012; Kujala, 2013). In their work, the authors compared button, swipe gesture and kinetic scrolling for grid (Lasch and Kujala, 2012) and linear lists (Kujala, 2013). While button and swipe gestures (i.e. unidirectional finger movement to trigger an action) allow the user to browse through pages with a fixed-number of items, kinetic scrolling allows for a more continuous browsing (i.e. dragging the view-port with finger movement) including scrolling rate control. Both studies concluded that kinetic scrolling performed generally worse and was more distracting than swipe gesture and touch-buttons.

Rydström et al. (Rydström et al., 2012) tested multiple widget types and two types of input methods (touch-screen or physical rotary control). They found that while both types of input method affected longitudinal control, touch-screen based interactions impacted lateral control to a greater extent. Moreover, the rotary control impacted on the performance to a lesser extent (better control of the vehicle and fewer off-road glances) for continuous adjustment tasks (e.g. radio, volume or list searching). One possible explanation of these results is that with a rotary control drivers don't have to physically reach the screen and/or rely on poor screen resolution for discriminating between targets. According to Kim and Song (Kim and Song, 2014) gesture-based interactions are more often worse than their classic touch button counterparts when used within an in-car set-up. Only the panning gesture was found to have a small impact on driving performance. In contrast, the flicking gesture (kinetic scrolling) or pinching were found to be very difficult to control. Similarly, Young et al. (Young et al., 2012) found that continuous use of kinetic scrolling when searching for music on an MP3 player significantly impaired driving performance. Finally, Kujala et al. (Kujala et al., 2013) concluded that text entry and kinetic scrolling are major sources of visual-manual distraction in the car.

The body of evidence concerning multi-touch gestures in general and kinetic scrolling in particular may be the opposite of what would be expected. Indeed, multi-touch gestures are supposed to require less accuracy in finger pointing than touch button interfaces and hence should be less difficult to use. In theory they should also reduce distraction. However, the opposite appears to be true. One possible explanation is that kinetic scrolling requires continuous visual-manual monitoring thereby decreasing the user's ability to interrupt the secondary task (for instance, as opposed to driving) and as a consequence this results in less safe behaviours (Chiang et al., 2004; Noy et al., 2004). These results emphasise the central role played by the interruptibility of a secondary task in the safe and effective completion of concurrent tasks (Burns et al., 2010). Indeed, the possibility to chunk a secondary task into multiple interaction steps allows for it to be interrupted and resumed when the primary task (in this case driving) necessitates it. Multi-touch gestures, as they require continuous visual-manual control may impair the ability of the user to interrupt the execution of the secondary task.

As pointed out by (Kujala, 2013) a page-per-page technique could improve the interruptibility of list browsing tasks; although kinetic scrolling might still be a better fit for long and ordered item lists. Indeed, kinetic scrolling allows users to skip large chunks of the list with only one movement, while the pager technique requires users to go through each page with a swipe or touch gesture. Additionally, as repetitive multi-touch gestures may cause fatigue in the user's wrist (Kim and Song, 2014) it is necessary to reduce the number of occurrences for these types of interactions. These results confirm that the way a user browses lists of items on multi-touch devices could still be improved. In

particular when the user may want to browse a long list by skipping non-relevant items.

In this work a driving simulator is used to assess three prototype tasks based on two envisioned use cases for the proposed i-Gear applications (McCall and Koenig, 2012; McCall et al., 2013). As we stated earlier, we envisioned two use-cases: one consisting in accepting or not sharing a ride with a friend, and the second one, consisting in selecting an alternative activity on the basis of the required duration of those activities. In this paper we want to assess the impact on driver distraction of those two use-cases if they were implemented in real-time and used while driving. We assess the two use-cases that are envisioned for the final application: As the "Alternative activities" use-case is potentially the most complex one, it is important to quantify its impact on drivers' distraction first against the "car sharing" use-case and then in further details under different implementations.

More precisely, we test three prototype tasks: the Help task implemented the "car sharing" use-case while the Browse and Filter ones are two different implementations of the "alternative activities" use-case (see Table 1). We present two implementations of the "alternative activities" use-case because it is potentially more disruptive and the flow of interactions could be organised in a one- or two-step way. To assess the impact of these prototype tasks on the driver's performance, three different types of metrics are used: the application usage performance (error rate and completion time), the telemetry of the car (lateral and longitudinal control) and gaze behaviour (number of off-the-road fixations and their durations). These performance metrics are compared to a baseline with two different methods: (1) between-trial (with and without application trials) and (2) within-trial (when dual-tasking and when driving only).

## 2. Material and methods

### 2.1. Participants

Twenty nine participants took part in this study (15 ♀/14 ♂). They were aged from 22 to 49 ( $30 \pm 6$ ,  $m \pm sd$ ). They had all held their driving license for at least four years ( $11.9 \pm 6.7$ ,  $m \pm sd$ ); this was done to ensure each participant possessed comparable minimum driving skills. Participants were drawn from the University staff and students population. There was no compensation offered for the participation. All participants signed an informed consent form complying with the University ethic committee guidelines (i.e., right to withdraw, usage of personal data).

### 2.2. Experimental design

Participants were placed in an empty rural-like simulated environment. The landscape was a flat textured surface with a two-lane 7 m wide road and grass on each side. Some additional decorations (22 bridges) were positioned along the route so as to improve the perception of speed.

In the driving task the participants drove on a straight track for 4 km. They were asked to follow a lead vehicle driving on the same lane without overtaking (known as a car-following task, see also Fig. 1, left). They were told to follow the lead vehicle and to keep a 2 s distance between themselves and the lead car for safety, although it was made clear to them that this constraint will not be enforced. We did not enforce this 2 s rule during the experiment in order to observe spontaneous distraction related compensation behaviours rather than provoking car collisions. The lead vehicle changed its speed 18 times during each trial at regular intervals. The speed was selected randomly under a uniform distribution centred on 45 km/h,  $\pm 15$  km/h.

**Table 1**

Presentation of the three different tasks and their description.

Use-case	Task	Description
Sharing a ride with a friend	Help	Touch buttons, "yes" or "no" answer
Choosing an alternative activity	Browse	Kinetic scrolling list, choose an activity among four categories of duration
	Filter	A slider allows to select an activity's duration, then a list containing only the relevant items is displayed



**Fig. 1.** Experimental task (left) and set-up (right): In half of the trials participants were to follow the lead car while they interacted with a docked smartphone in the driving simulator.

A smartphone task was triggered on every even count of the lead car's speed change. To prevent the participants from anticipating this event, the application trigger was placed randomly  $\pm 100$  m around the place where the speed change occurs. There were in total nine application triggers per trial resulting in each of the three prototype tasks being triggered three times per trial. The order in which the smartphone task appeared was randomised for each trial. Each smartphone task was first signalled by a visual and auditory notification, the title of the notification informed the participant about the nature of the task and the alternative action that they had to select (the "right" answer). Following this, another screen displayed the task's content with different alternatives.

### 2.3. Application design

We tested three mobile tasks on a smartphone. The Help task was implemented with a simple yes/no touch button interface (Fig. 2, left). The notification prior to the task indicated if the participant should choose "yes" or "no". The second use-case was implemented via two task prototypes (Fig. 2, middle and right). The Browse task provides a list of alternative activities ordered by duration two activities for each four duration categories, which are browsed using the kinetic scrolling technique (Fig. 2, middle). Finally, the Filter task, proposes first a slider interface for selecting a range of desired durations for the alternative activity. Once the range is selected the prototype shows only the two relevant activities (Fig. 2, top and bottom right). Each time category was defined with chunks of 30 min along the slider: positioning the cursor from 15 to 45 min resulted in the selection of the 30 min category, from 45 to 75 resulted in the selection of the 60 min category, etc. The participants were familiarised with the user interface of each task before the data collection phase. For all the prototype tasks, the prior notification indicates both the nature of the oncoming task and the item that should be selected (for alternative activities, only the duration category was specified, participants were free to choose any of the two activities in the specified category). This design has been chosen so as to limit cognitive distraction due to decision making which is beyond the scope of this work.

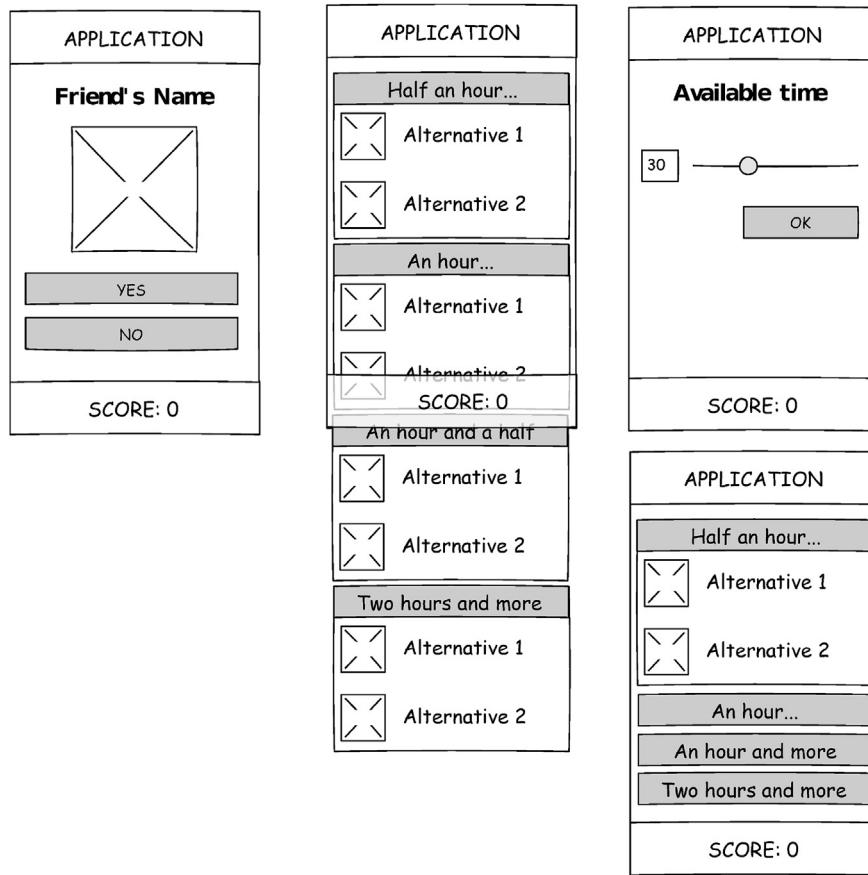
### 2.4. Driving simulator and cockpit

In this study we used a low-cost driving simulator set-up (see also Jamson and Jamson (Jamson and Jamson, 2010)) which uses standard driving game controllers, a screen and a PC. Participants were placed in front of a Panasonic plasma screen of 165 cm diagonal, full HD resolution. Car controls were provided by a Fanatec Porsche 911 GT2 steering wheel and pedal set. The simulator used an automatic transmission gear box. The 3D simulation engine used was OpenDS 1.0 (Math et al., 2013) (modified in order to obtain telemetry data in real time) and the overall experimental script (events and mobile application) was orchestrated by the DriveLab testing platform (Louveton et al., 2013; Avanesov et al., 2012). The DriveLab platform supports the scripting and logging of events (e.g. from/to mobile devices) and car telemetry data. Real-time data captured from the 3D engine and the application are stored in a database and are timestamped, this allows for easy synchronisation and off-line data analysis. Additionally, eye-tracking data was captured with SMI Eye Tracking Glasses.

### 2.5. Mobile device and implementation

The mobile application prototype was implemented on a Huawei Ascend G330 smartphone with 10.2 cm TFT screen (480×800 resolution) and 512 Mb of RAM memory running Android 4.0. One critical aspect is that the studied prototypes should have a high face validity outside of the laboratory setting. For this reason our mobile application prototypes were developed in HTML5 using the JQuery mobile GUI framework which is both a convenient prototyping tool and a plausible framework for developing market level applications.<sup>1</sup> We used Firefox for Android and the full-screen add-on in order to run the mobile application prototypes. The smartphone was placed on the right side of the steering wheel within a standard dock (suction cup) (see Fig. 1, right panel). The orientation of the smartphone's screen was adjustable but not the position of the dock relative to the steering wheel.

<sup>1</sup> See, for example, [www.jqmgallery.com](http://www.jqmgallery.com).



**Fig. 2.** Schematic representation of the task prototypes: Help (left) task that requires a binary decision (yes/no) from the user, Browse (middle) that provides four time categories with two selections for each and Filter (right) requires the user to specify an amount of spare time (top-right) and then to choose one of two relevant selections (bottom).

## 2.6. Procedure

Participants first had to complete a profile questionnaire and an informed consent form. Next they were placed in the driving simulator; they were able to adjust the seat and pedal position if required. They then read an instruction sheet. Next the test administrator demonstrated each of the mobile application tasks and gave the participants the opportunity to perform one test action. Finally, the eye-tracker was calibrated (using three points calibration).

In order to normalise training effects during the study, all participants completed a familiarisation task. During this task they drove along a 1 km road within the simulation environment. The purpose of this task was to allow people to familiarise themselves with the nature of the controls (e.g. steering wheel and pedals) and to become accustomed to the seating position and field of view within the simulator. After completing the familiarisation task they took part in the data collection phases that consisted of four experimental trials: two with the application running on the smartphone and two without. The order of the four trials was randomised for each participant. After completion of the trials participants were asked to complete a questionnaire.

This small questionnaire (i.e., eight items) was designed to allow for an assessment of subjective aspects such as perceived ease of use and the strategies used by the drivers. The five first questions were Likert scale type while the three last ones were open-ended questions. The likert scale items were asking about: the difficulty of the driving task without (1) and with (2) the application, and the difficulty of the three tasks: Help (3), Filter (4) and Browse (5). Each

scale ranged over 5 steps with 1 indicating very easy and 5 very hard. Finally, the open-ended questions let the participants commenting on the strategies they used for accomplishing the driving task with and without the application, and also any potential comments on the experiment.

## 2.7. Data analysis

Application usage performance was analysed in terms of task completion time and error rate (i.e., congruence between the user's answer and the notification requirement). Concerning driving performance, we analysed both longitudinal and lateral control. Longitudinal control was operationalised through speed and CG Headway (CG standing for centre of gravity). The CG Headway was calculated as the distance between the two cars' geometric centre (also see SAE J2944 ([Society of Automotive Eng \(2013\)](#))). For lateral control we analysed lane position and the Standard Deviation of Lane Position (SDLP hereafter). It was calculated using the unbiased estimation of standard deviation applied on the lane position data (i.e. distance from the lane and car's geometric centre). See also SDLP option A of SAE J2944 ([Society of Automotive Eng \(2013\)](#)). Driving performance is compared to baseline either between-trial (trials with and without the application) or within-trial (interacting or not with the application).

For the between-trial analysis, data was averaged across trials before comparing between with and without application trials. Concerning the within-trial comparison, data was averaged across chunk type (not using the application, help, browse or filter tasks) before comparing the means of the different categories. Statistical

tests used were all accounting for repeated-measures design and were either parametric (*t*-test for two-samples, or ANOVA for multiple samples) or non-parametric (wilcoxon and friedman tests respectively) when normality was not achieved. We used  $\log_{10}$  transformation in case of right-sided skewness of the distribution in order to reach normality. Multiple comparisons post-hoc analyses were conducted using either Tukey contrasts after a repeated measure ANOVA or pair-wise wilcoxon test comparisons with Bonferroni correction after a friedman test. Data analysis was carried out with Pandas (0.15) library (McKinney, 2011) for Python (2.7) and R (3.1.2) (R Core Team, 2013) with some extra libraries such as *nlme* (Pinheiro et al., 2015) and *multcomp* (Hothorn et al., 2008).

Gaze data was coded in a frame-by-frame fashion, using a region of interest for detecting whether or not the participant was looking at the device while driving. We coded the fixations on the device screen (i.e. gaze located in the region of interest), in which we did not include the eye-movement duration. The gaze data was then synchronised with the telemetry and application data using a signal sent by the DriveLab server when the participant started and finished each trial. We were therefore able to relate specific gaze events with driving performance and mobile interaction. We analysed the mean single-fixation duration, the number of fixations and the cumulative fixation time (total fixation time for one occurrence of the task).

### 3. Results

#### 3.1. Subjective responses

Data from the questionnaire likert responses indicated that the participants found the driving task more difficult when they were using the application on the smartphone (when compared to not using it). On a difficulty scale ranging from one to five, the median score for the without-application trials was 2 while the median for with-application trials was 3. The per-task analysis demonstrates that the Help task was judged to be the easiest with a most frequently observed score of 1 and a median score of 1. The two other tasks appeared to be more difficult. Indeed, Browse showed a median score of 2 and Filter a median score of 3.

Tables 2 and 3 summarise the comments provided by participants. The comments indicate that the introduction of the application into the scenario had an impact on their awareness of the driving task, for example, in such aspects as "application related behaviour" or "holding the phone". However, the participants seemed to be much more aware of other aspects of the task when being under the mobile device condition. Indeed, they explicitly indicated using the car controls more frequently and there were two extra comments relating to their behaviour relative to the lead car.

#### 3.2. Application usage performance

Overall 90.4% of all tasks were successfully completed (469/522). The Filter task appeared to have less successful occurrences:

**Table 2**

Categories of comments expressed by participants about executing a car-following task without using the mobile application.

Without application comments	No. of comments
Behaviour relative to lead car (incl. Focus on lead car)	5
Behaviour relative to environment	2
Maintaining speed	1
Less braking due to using accelerator	1
No strategy	2
Adjusting hand relative to mobile device	1
Total	12

**Table 3**

Categories of comments expressed by participants about executing a car-following task while using the mobile application.

With application comments	No. of comments
Behaviour relative to lead car (incl. focus on lead car)	7
Behaviour relative to environment	1
Use of car controls	5
Application related behaviour	6
Time	1
Keeping the lead straight (not fully legible)	1
Personal behaviour e.g. holding phone or adjusting cockpit	2
General – focus	1
Simulator issues	3
Total comments	27

146 out of 174 (83.9%) against 161 for Help (92.5%) and 165 for Browse (94.8%) tasks. Also, the highest error rate has been found in Filter task (12%), compared to Browse (5%) and Help (3.4%).

The average completion time (for correct responses only, see Fig. 4) varied significantly with the type of task (Friedman rank sum test:  $\chi^2(2) = 28.86; p < .001$ ). Post-hoc comparisons using a Wilcoxon paired tests with a Bonferroni correction demonstrated that each task differed significantly in terms of completion time ( $p < .001$ ). Indeed, the Filter task took the longest to complete (15.25 s, SD = 5.73) followed by the Browse (12.18 s, SD = 4.39) and Help ones (8.72 s, SD = 2.43).

#### 3.3. Driving performance

##### 3.3.1. Between-trial comparisons

When comparing the average performance for with vs without application trials (see also Fig. 3) we found that participants significantly increased their CG Headway when they were using the smartphone; as shown by a paired *t*-test,  $t(57) = 6.0453, p < .001$ . The mean CG headway distance for with application trials was 38.27 m (SD = 26.07), compared to 25.19 m (SD = 10.63) in without application trials. The SDLP indicator was also statistically different ( $t(57) = -77.9, p < .001$ ) with a more variable lateral control when participants were using the smartphone (0.32 m, SD = 0.14) than when they were not (0.27 m, SD = 0.07).

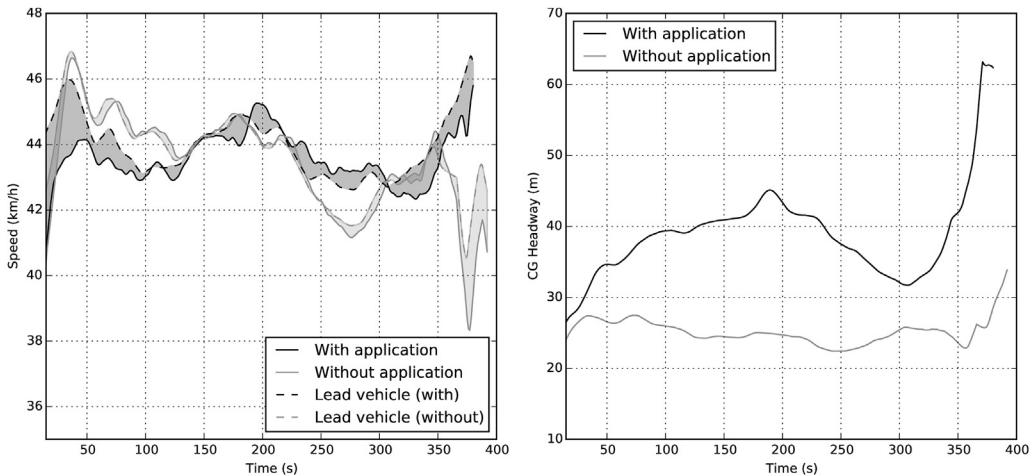
We found no significant differences for the average speed ( $p = .67$ ) and the average lateral position ( $p = .34$ ). Indeed, there were practically no differences in average speed (with-application trials: 43.79 km/h, SD = 2.49; without-application trials: 43.98 km/h, SD = 2.91) and in average lateral position from the centre of the lane (with-application trials: 0.14 m, SD = 0.20; without-application trials: 0.12 m, SD = 0.18).

##### 3.3.2. Within-trial comparisons

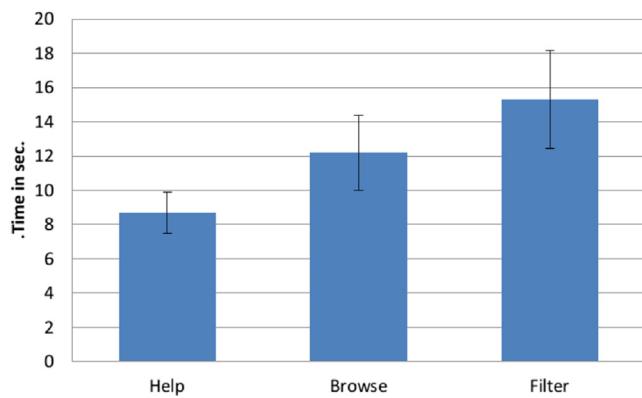
In this analysis we compare the different driving performance metrics in four experimental conditions (Baseline, Help, Browse and Filter) within the trials where the participants used the application (see also Fig. 5).

We found near significant differences in CG Headway ( $p = .06$ ) and Lateral position ( $p = .08$ ). The CG Headway for the four conditions was 35.9 m in Baseline condition (SD = 25.7), 36.9 m in Help (SD = 27.5), 41.3 m in Browse (SD = 30.5) and 44.5 m in Filter (SD = 28.2). The lateral deviation values were 0.11 m for Baseline condition (SD = 0.20), 0.15 m for Help task (SD = 0.22), 0.14 m for Browse task (SD = 0.20) and 0.22 m for Filter task (SD = 0.26).

However we found a significant difference for average speed ( $F(3,228) = 2.89, p < .05$ ). The post-hoc analysis (Tukey's contrasts) showed a significant difference between Baseline and Help task ( $p < .05$ ): participants drove slightly faster in the Help condition than in the Baseline one. Overall, participants drove at an average



**Fig. 3.** The left panel shows a comparison of with (black) and without (gray) application trials: the synchronisation between participants' speed (plain line) and lead vehicle's speed (dashed line) is higher in the without application condition (smaller surface area is better). The right panel shows the cumulative difference in participants'/lead-vehicle's speeds synchronisation: participants cumulated a bigger CG Headway over time in the with application condition.



**Fig. 4.** Completion time varied significantly across the different tasks, the Filter task being the longest to complete. Such results have to be compared to fixation time results to make further sense.

speed of 43.4 km/h ( $SD = 2.9$ ) in Baseline condition, 45.7 km/h in Help ( $SD = 5.1$ ), 43.9 km/h in Browse ( $SD = 4.7$ ) and 44.1 km/h in Filter ( $SD = 4.6$ ).

We also found a significant difference for SDLP ( $F(3,228) = 4.16, p < .01$ ). The post-hoc comparison revealed that lateral control was significantly more variable ( $p < .01$ ) in Filter ( $0.34, SD = 0.18$ ) than in Help ( $0.27, SD = 0.14$ ) conditions. We also found a near significance difference ( $p = .06$ ) between Browse

( $0.28, SD = 0.13$ ) and Filter ( $0.34, SD = 0.18$ ) suggesting more variable lateral control in this latter condition. We found no significant difference involving the Baseline condition ( $0.29, SD = 0.11$ ).

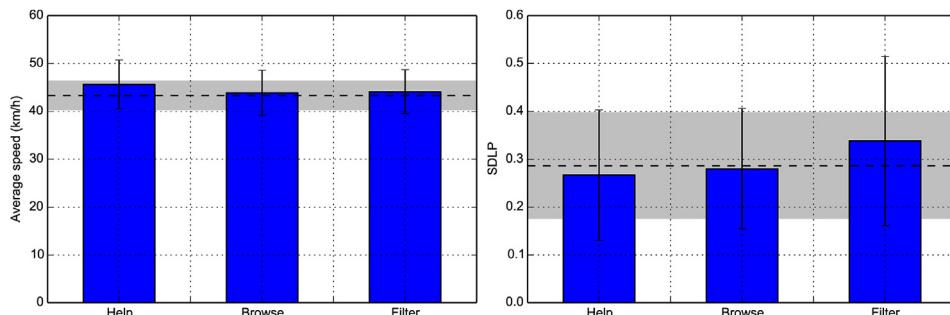
### 3.4. Gaze data analysis

#### 3.4.1. Number of fixations

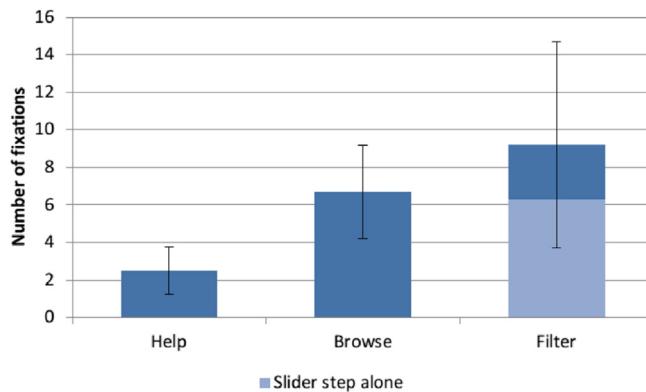
Gaze data analysis was carried out using the Friedman test and focussed on comparing the number of fixations towards the device instead of the road (Fig. 6). This was shown to be affected by the application type ( $\chi^2(2) = 74.34; p < .001$ ). A pair-wise Wilcoxon rank sum test was performed to compare those distributions (Bonferroni correction) and we found all comparisons to be significant ( $p < .001$ ). More precisely, participants glanced more often to the Filter task screen (9.22 glances average per instance of the task,  $SD = 11$ ), followed by the Browse task (6.68 glances,  $SD = 5.04$ ) and Help task (2.49 glances,  $SD = 2.54$ ).

#### 3.4.2. Cumulative fixation time

The cumulative fixation time on task completion was shown to be affected by the application type ( $\chi^2(2) = 90.82; p < .001$ ). A pair-wise Wilcoxon rank sum test was performed to compare those distributions (Bonferroni correction) and significant differences were found across all the proposed tasks ( $p < .001$ ). For instance, participants spent more time glancing at the Filter task screen (5.1 s on average per instance,  $SD = 3.9$ ), followed by the Browse task (3.8 s,  $SD = 2.7$ ) and Help task (1.7 s,  $SD = 2$ ).



**Fig. 5.** Average speed (left) and SDLP (right) are represented in the four experimental conditions (the dashed line represents the baseline condition and the gray area the standard deviation).



**Fig. 6.** The number of fixations varied significantly between Help and Filter but not between Browse and Filter. It is worth mentioning that most of the fixations in the Filter task are due to the slider-step. The number of fixations dedicated to the slider-step are comparable to those of the complete Browse task.

#### 3.4.3. Mean single fixation time

Mean single fixation time also varied significantly across conditions ( $\chi^2(2) = 9.57; p < .01$ ). The post-hoc analysis revealed a significant difference between Help and Filter ( $p < .01$ ) and Browse and Filter ( $p < .05$ ): the Filter task being the one that required the longest mean single fixations (0.83 s,  $SD = 0.28$ ) followed by Browse (0.72,  $SD = 0.28$ ) and Help (0.7,  $SD = 0.27$ ).

## 4. Discussion

The goal of this study was to assess the impact on driving of three prototypes based on the two envisioned use-cases for a smart-phone mobility application. The purpose of the application was to present people with a series of alternative mobility choices. This was carried out by comparing the use of the application to a baseline condition. Firstly, the between-trial performance was compared (with vs without application trials). Then within-trial data (users interacting with the device vs not interacting with the device) was also compared. Finally, performance metrics from each of the three prototype tasks was compared.

#### 4.1. Comparison to the baseline performance

The between- and within-trial baseline comparisons emphasised different patterns of degradation in driving performance. The between-trial comparison illustrated a moderate effect of using the application on the SDLP and a strong effect on the CG Headway. Indeed, analysis revealed the participants de-synchronise their displacement with respect to the lead car. This phenomenon has already been observed in prior studies (Salvucci et al., 2007; Young et al., 2012; Lasch and Kujala, 2012), and is a sensible behaviour as it increases the margin of tolerance should the driver need to brake suddenly. Additionally, the absence of statistical difference with respect to average speed and lateral deviation indicates that averaging speed and lateral position over the trial unfolding smooths the differences between the two types of trial. In contrast, CG Headway and SDLP cumulate the variations over time making those metrics more sensitive in between-trial analysis. In contrast, the within-trial comparison found a moderate effect (yet statistically significant) on SDLP and lateral deviation (both increased when the user interacted with application). We also found near-significance differences for speed and CG Headway which were also moderate in amplitude. This comparison illustrates that the chunked comparison is less sensitive to compensation (compared to the between-trial comparison); although it could lack statistical power if the effects are moderate in amplitude.

#### 4.2. Comparison of the three prototype tasks

Completion time and success rate analyses suggested that the Filter task was the most difficult as it was more error prone and took longer to complete successfully. The analysis also found that the Browse task took much longer to complete than the Help task. When comparing the effect of the different tasks on driving performance the results point to a mild impairment in performance, specifically with respect to the SDLP metrics. Indeed, the Filter task induced a more variable lateral control than other tasks, while no differences were found between Help and Browse. Analysis of gaze data was more sensitive than driving performance. We analysed single-fixation duration, cumulative fixation time and the number of fixations associated with each task. The results indicate that the cumulative fixation time was the highest for the Filter task followed by the Browse and Help tasks. This increase in fixation time is mainly due to a higher number of fixations. Indeed, the number of fixations was statistically different for Help compared to the Browse and Filter tasks. There were also differences in terms of mean single-fixation duration, although of a smaller importance. The longer cumulative fixation time for the Filter task cannot be explained solely by the two-step nature of the task. Indeed, participants spent on average 75% of the total completion time on the slider step, and total fixation time on this step represents 35% of the task's completion time (compared to 13% for the selection step). The participants tried to maintain a low single-fixation duration regardless of the task they perform. In essence this meant that the participants altered their number of fixations such that they viewed the interface in short bursts distributed over time rather than in larger or single chunks. In terms of safety, shorter fixations are better as they allow for improved chunkability of visual time-sharing between the two tasks (Burns et al., 2010) (e.g. driving vs interacting).

#### 4.3. Implications for information browsing and interruptibility

Although the results illustrate that the Help task was arguably the easiest one, they also illustrate that both Browse and Filter had their own level of difficulty within a dual-tasking context. Indeed, while the Browse task required browsing through a list using a kinetic scrolling, the Filter task replaced this browsing process with a multi-step approach (using a slider widget to filter the content of the list). In former literature (Young et al., 2012; Kujala et al., 2013; Lasch and Kujala, 2012), kinetic scrolling has been emphasised as lacking of interruptibility. We hypothesised that the lack of interruptibility of the kinetic scrolling method was due to the lower predictability of view-port displacement. In this respect, we assumed the multi-step approach of the Filter task to be more interruptible than the Browse one. This approach uses a cursor motion which has a more predictable control (i.e. one-to-one match with finger movement and stays on the area if released) than kinetic scrolling. This could have allowed the user to stop and resume the interaction with the device as often as needed.

However, the results showed that the Filter task was more difficult than the Browse one (i.e., using kinetic scrolling). Most of the time spent looking at the device was due to the specific filtering step using the slider. Indeed, the slider widget also requires a high accuracy in finger pointing movement. This high demand for accuracy in finger pointing probably made the participants reluctant to release the cursor during the filtering step. However, deeper analysis shows the benefit of a filtering step should improve the selection process as the fixation time dedicated to the selection step represented only 27% of total fixation time. It is also worth mentioning that the time spent glancing at the device on the filtering step is comparable to that of the complete Browse task. As

kinetic scrolling requires less finger pointing accuracy (Fitts, 1954) than a slider, it is thus surprising that the fixation metrics for both situations are equivalent.

On one hand, the fact that kinetic scrolling exhibited poor performances similar to those of the slider widget confirms former literature pointing to kinetic scrolling being a major source of visual-manual distraction while driving (Kujala et al., 2013), even with a small list like in our case (eight items spread across four categories). On the other hand, this also demonstrates that Fitts' law (Fitts, 1954) is not sufficient for predicting performance of multi-touch interfaces in dual-task context as it does not include interruptibility issues. However, even if the slider filtering method has been shown to be the most distracting prototype according to our metrics, this seems to be mainly due to the slider step (for a break-down of filtering and selecting steps contribution in eyes fixations, see Figs. 6 and 7). This suggests that the multi-step approach is worthy of further research; particularly if the filtering step were less demanding in terms of visual-manual control.

#### 4.4. Limitations of the study

Potential limitations of this study include the lack of validity of the driving simulator and the low difficulty level of the driving task. For instance, we used a low-cost driving simulator (both in terms of the cockpit and 3D engine) and it has been shown that low-cost simulators can degrade the accuracy of self-displacement or inter-vehicular distance perception (Kemeny and Panerai, 2003). However, the same authors emphasised that low-cost simulators can be useful for simple driving scenarios or dashboard evaluation. Additionally, it has been shown that low-cost simulators are useful for early prototyping of IVIS, in particular when the analysis focuses on speed control and completion time of the secondary task (Green, 2004; Jamson and Jamson, 2010). Finally, the driving task used in this study was relatively simple, and consisted of a rural, straight-road with only one additional vehicle. Such a driving condition could be considered as placing a low mental workload on the driver (Paxion et al., 2014). Hence replicating the study with different driving scenarios with a range of imposed mental workloads would allow for a higher generalisation of the results.

#### 4.5. Future research

Two research lines are suggested by this work. The first consists in gaining a better understanding of why kinetic scrolling performs poorly and how to improve it (as it could apply to maps or long

lists). Improving the multi-touch browsing process would include the possibility to perform a two-scale search: (1) a large-scale and non-accurate search for filtering non relevant information (2) a small-scale and accurate one to select the desired options. The second line of research is about how to improve the interruptibility of a task by splitting it into several steps. We have seen that replacing browsing of information with a filtering method reduces the time spent on the selection of an alternative. This raises the question of how many steps can be used in order to improve interruptibility while simultaneously maintaining task completion time within a reasonable range.

## 5. Conclusions

In this study we tested three prototype tasks that implemented two possible use-cases for a mobility application. We found that it is not strictly necessary to perform a between-trial baseline comparison, provided that the effects of the experimental manipulations are large enough to be detected in the within-trial comparison. When comparing the different tasks to each other, we found that the Filter task was the most demanding in terms of visual-manual control. Further analysis showed that most of the time spent looking at the device was due to the filter step requiring the control of a slider widget. The amount of gaze duration toward the device during the filter step was comparable with one of the whole Browse task. While manipulating a slider is demanding in terms of visual-manual control (i.e., small target manipulation), it is interesting to note that the kinetic scrolling search performed in a comparable way, reproducing former results from literature on list browsing. The results suggest that it is important to better understand and improve browsing of long documents (lists or maps) and how to design multi-step interactions.

## Acknowledgements

The i-Gear project was funded by the Fond National de la Recherche (FNR), Luxembourg (Grant number: 11/IS/1204159). We also thank other colleagues and students from the i-Gear team.

## References

Alm, H., Nilsson, L., 1995. The effects of a mobile telephone task on driver behaviour in a car following situation. *Accid. Anal. Prev.* 27 (5), 707–715.

Avanesov, T., Louveton, N., McCall, R., Koenig, V., Kracheel, M., 2012. Towards a simple city driving simulator based on speed dreams and OSM. In: *Automotive UI in Adjunct Proceedings – Work in Progress Paper*.

Burns, P., Harbluk, J., Foley, J.P., Angell, L., 2010. The importance of task duration and related measures in assessing the distraction potential of in-vehicle tasks. In: *Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, pp. 12–19.

Chiang, D.P., Brooks, A.M., Weir, D.H., 2004. On the highway measures of driver glance behavior with an example automobile navigation system. *Appl. Ergon.* 35 (3), 215–223.

Fishman, T., 2012. *Digital-age Transportation: the Future of Urban Mobility*. Technical report. Deloitte University Press.

Fitts, P.M., 1954. The information capacity of the human motor system in controlling the amplitude of movement. *J. Exp. Psychol.* 47 (6), 381.

Green, P., 2004. *Driver Distraction, Telematics Design, and Workload Managers: Safety Issues and Solutions*. Technical report. SAE International.

Hick, W.E., 1952. On the rate of gain of information. *Q. J. Exp. Psychol.* 4 (1), 11–26.

Horberry, T., Anderson, J., Regan, M.A., Triggs, T.J., Brown, J., 2006. Driver distraction: the effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accid. Anal. Prev.* 38 (1), 185–191.

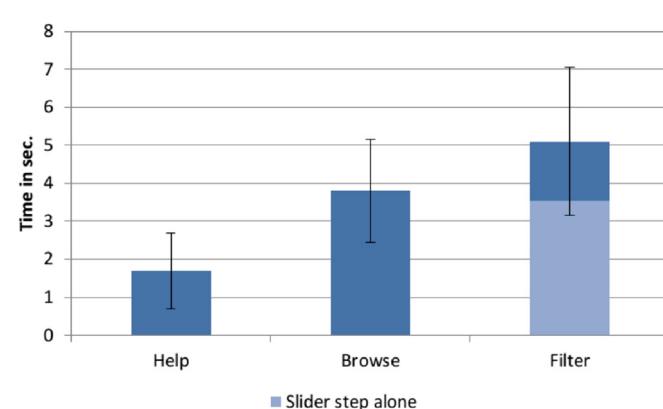
Hothorn, T., Bretz, F., Westfall, P., 2008. Simultaneous inference in general parametric models. *Biom. J.* 50 (3), 346–363.

Jamson, S.L.S., Jamson, A.H., 2010. The validity of a low-cost simulator for the assessment of the effects of in-vehicle information systems. *Saf. Sci.* 48 (10), 1477–1483.

Kemeny, A., Panerai, F., 2003. Evaluating perception in driving simulation experiments. *Trends Cogn. Sci.* 7 (1), 31–37.

Kim, H., Song, H., 2014. Evaluation of the safety and usability of touch gestures in operating in-vehicle information systems with visual occlusion. *Appl. Ergon.* 45

**Fig. 7.** Fixation time analysis revealed that the cumulative time spent looking at the device increased significantly across the different tasks. The fixation time overhead for the Filter task was due to the slider-step which required almost as much fixation as the Browse task.



(3), 789–798.

Kujala, T., 2009. Efficiency of visual time-sharing behavior: the effects of menu structure on poi search tasks while driving. In: Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications. ACM, pp. 63–70.

Kujala, T., 2013. Browsing the information highway while driving: three in-vehicle touch screen scrolling methods and driver distraction. *Personal Ubiquitous Comput.* 17 (5), 815–823.

Kujala, T., Silvennoinen, J., Lasch, A., 2013. Visual-manual in-car tasks decomposed: text entry and kinetic scrolling as the main sources of visual distraction. In: Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. ACM, pp. 82–89.

Lasch, A., Kujala, T., 2012. Designing browsing for in-car music player: effects of touch screen scrolling techniques, items per page and screen orientation on driver distraction. In: Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. ACM, pp. 41–48.

Louveton, N., McCall, R., Avanesov, T., Koenig, V., Kracheel, M., 2013. Assessing in-vehicle information systems application in the car: a versatile tool and unified testing platform. In: Automotive UI Proceedings.

Math, R., Mahr, A., Moniri, M.M., Müller, C., 2013. Opends: a new open-source driving simulator for research. In: Gmm-fachbericht-ame 2013.

McCall, R., Koenig, V., 2012. Gaming concepts and incentives to change driver behaviour. In: Ad Hoc Networking Workshop (Med-hoc-net), 2012 the 11th Annual Mediterranean. IEEE, pp. 146–151.

McCall, R., Koenig, V., Kracheel, M., 2013. Using gamification and metaphor to design a mobility platform for commuters. *Int. J. Mob. Hum. Comput. Interact.* (IJMHCI) 5 (1), 1–15.

McKinney, W., 2011. Pandas: Powerful python Data Analysis Toolkit.

Noy, Y.I., Lemoine, T.L., Klachan, C., Burns, P.C., 2004. Task interruptability and duration as measures of visual distraction. *Appl. Ergon.* 35 (3), 207–213.

Paxion, J., Galy, E., Berthelon, C., 2014. Mental workload and driving. *Front. Psychol.* 5 (1344).

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Core Team, R., 2015. Nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-119.

R Core Team, 2013. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rydström, A., Broström, R., Bengtsson, P., 2012. A comparison of two contemporary types of in-car multifunctional interfaces. *Appl. Ergon.* 43 (3), 507–514.

Salvucci, D.D., Markley, D., Zuber, M., Brumby, D.P., 2007. Ipod distraction: effects of portable music-player use on driver performance. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, pp. 243–250.

Society of Automotive Engineers (Proposed draft 2013). Driving Performance Definitions. SAE J2944, Surface Vehicle Recommended Practice.

Treffner, P.J., Barrett, R., 2004. Hands-free mobile phone speech while driving degrades coordination and control. *Transp. Res. Part F Traffic Psychol. Behav.* 7 (4–5), 229–246.

Young, K., Regan, M., Hammer, M., 2007. Driver Distraction: a Review of the Literature. Distracted driving. Australasian College of Road Safety, Sydney, NSW, pp. 379–405.

Young, K.L., Mitsopoulos-Rubens, E., Rudin-Brown, C.M., Lenné, M.G., 2012. The effects of using a portable music player on simulated driving performance and task-sharing strategies. *Appl. Ergon.* 43 (4), 738–746.