

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

Author(s): Kujala, Tuomo; Grahn, Hilikka

Title: Visual distraction effects of in-car text entry methods : Comparing keyboard, handwriting and voice recognition

Year: 2017

Version:

Please cite the original version:

Kujala, T., & Grahn, H. (2017). Visual distraction effects of in-car text entry methods : Comparing keyboard, handwriting and voice recognition. In S. Boll, B. Pfleging, B. Donmez, I. Politis, & D. Large (Eds.), *AutomotiveUI '17 : Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 1-10). ACM. <https://doi.org/10.1145/3122986.3122987>

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

Visual Distraction Effects of In-Car Text Entry Methods – Comparing Keyboard, Handwriting and Voice Recognition

Tuomo Kujala
University of Jyväskylä
Jyväskylä, Finland
tuomo.kujala@jyu.fi

Hilkka Grahn
University of Jyväskylä
Jyväskylä, Finland
hilkka.grahn@jyu.fi

ABSTRACT

Three text entry methods were compared in a driving simulator study with 17 participants. Ninety-seven drivers' occlusion distance (OD) data mapped on the test routes was used as a baseline to evaluate the methods' visual distraction potential. Only the voice recognition-based text entry tasks passed the set verification criteria. Handwriting tasks were experienced as the most demanding and the voice recognition tasks as the least demanding. An individual in-car glance length preference was found, but against expectations, drivers' ODs did not correlate with in-car glance lengths or visual short-term memory capacity. The handwriting method was further studied with 24 participants with instructions and practice on writing eyes-on-road. The practice did not affect the test results. The findings suggest that handwriting could be visually less demanding than touch screen typing but the reliability of character recognition should be improved or the driver well-experienced with the method to minimize its distraction potential.

Author Keywords

Driver distraction; visual demand; visual occlusion; occlusion distance; text entry methods; visual short-term memory; Visual Patterns Test.

CCS Concepts

• Human-centered computing~HCI design and evaluation methods • Human-centered computing~Graphical user interfaces • Human-centered computing~Interaction devices

INTRODUCTION

According to several studies, text entry with a touch screen keyboard is among the most visually distracting in-car tasks for the driver (e.g., [17,28,24]). Yet, it seems that many drivers are willing to take the risk, as it seems that short messaging with a smartphone is among the most popular in-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

car activities (e.g., [11]). Many in-car activities, that can support the primary task of driving, such as destination entry (way-finding) and music search (entertainment for keeping alert), may also require text entry. For these reasons, there is a need for visually less demanding in-car text entry methods than the touch screen keyboard (e.g., [29]).

In this study, three different in-car text entry methods were compared: touch screen keyboard, handwriting and voice recognition. A voice recognition-based text entry has been shown to be significantly less distracting than a keyboard text entry in several controlled studies (e.g., [9,10]). However, as pointed out by Reimer and Mehler [22], against common belief, also the voice-guided systems typically include some visual-manual interactions, which may be distracting. Handwriting on a touch screen is a rather new method for the automotive context, and it appears that there is not yet much published research concerned with this method. For example, Kern et al. [12] studied handwritten text in the automotive context, but comparative distraction testing of handwriting as a text entry method seems to be lacking. The advantage of handwriting is that it may enable the driver to keep eyes on the road while writing especially if the system gives audio feedback to the driver, that is, repeats the written letters out loud.

According to Foley, Young, Angell and Domeyer [5, p. 62], “visual distraction is any glance that competes with activities necessary for safe driving”. The definition of visual distraction by Foley et al. [5] is incomplete, as it does not define the “activities necessary for safe driving”. This incompleteness places challenges for the operationalization of visual distraction. According to the study by Kircher and Ahlstrom [13] there is minimum required attention for each driving situation that can be fulfilled by different visual sampling patterns off road. This suggests that not all off-road glances are equally distracting but the timing of an off-road glance plays a critical role in visual distraction. A distracting off-road glance can be interpreted as a calibration failure between the (momentary) visual demands of driving and the individual preference for an off-road glance length, following the task-capability interface model by Fuller [6]. Here, we refer to visual distraction, in short, as a calibration failure between driving task's visual demand and the driver's off-road glance length.

Following these lines of thought, a novel distraction testing method, introduced by Kujala and Mäkelä [15] was used in our study to evaluate and compare the visual distraction potential of the three text entry methods. The testing method has been used previously to study distraction potential of audio-visual route guidance (see [14]).

The testing method is based on 97 drivers' preferred occlusion distance (OD) data mapped on the test routes [16]. The concept of occlusion distance refers to a distance that is traveled during the occluded period, that is, the distance that a driver feels comfortable to drive without visual information while concentrating on the driving task. In the testing, the median ODs of the 97-driver sample for each 1-by-1-meter test route point are utilized as a baseline for acceptable in-car glance lengths (distances, to be exact). These are labeled as *green* in-car glances. The in-car glance distances exceeding the 85th percentile of the 97-driver sample at a road point are considered as calibration failures (following Fuller, [6]) and are labeled as *red* in-car glances. A red glance suggests that the in-car task has (momentarily) caught the driver's visual attention for a longer period of time than what a great majority of drivers would not prefer to glance off road when focusing on driving at that route point.

The testing method also strives to take the drivers' individual off-road glance length preferences into account as previously has been studied that these can significantly affect the results of the distraction testing (e.g., [2]). In order to analyze the reliability and validity of the test results, we studied the test participants' individual preferences for in-car glance distances and ODs and validate the comparability of the latter with the OD distribution of the 97-driver sample. In addition, we were interested to see if the OD preference could be explained by a capability-related measure of visual short-term memory capacity (Visual Patterns Test; [4]) and if the experiences of task demands are in line with the objective distraction metrics. The specific research questions for the distraction testing were:

1. Do the studied in-car tasks pass the used verification criteria set by Kujala et al. [14]?
2. Are there significant differences in the visual distraction potential between text entry by keyboard, handwriting and voice recognition?
3. Do the drivers experience different levels of task workload between the text entry methods?

For method validation, the research questions were:

4. Are the individually preferred ODs of the test participants comparable to the ODs of the baseline sample of 97 drivers [16]?
5. Is there an individual preference threshold for the in-car glance distances across the tasks?
6. Do the driver's OD's correlate with their preferred in-car glance distances across the tasks?

7. Do the drivers' ODs correlate with their Visual Patterns Test scores [4]?

EXPERIMENT 1 - COMPARATIVE DISTRACTION TEST

The experimental design of the distraction testing was within-subjects (one IV with three levels), the independent variable being the text entry method (keyboard, handwriting, voice recognition).

Participants

The NHTSA [20] recommendations on the driver sample for testing distraction of in-vehicle electronic devices were followed as closely as possible. The participants were recruited via university's mailing lists. In total 17 participants finished the experiment, twelve males and five females. Seven female participants had to quit the experiment because of symptoms of simulator sickness. Three of them were able to complete the occlusion trial and the Visual Patterns Test, and thus, the correlation tests between these include 20 participants.

The age of the participants varied from 20 to 63 years, mean age being 34.4 years ($SD = 12.2$). Five of the participants were 18 to 24 years old, six 25 to 39 years old, four 40 to 54 years old and two were older than 55 years. All participants had a valid driver's license and all of them drove at least 5 000 kilometers per year. The total kilometers driven per year varied from 7 000 to 30 000 and with a mean of 15 352 kilometers ($SD = 7 526$) per year. The driving experience varied from two to 45 years, with a mean of 16.8 ($SD = 12.0$) years. All participants had normal vision. The experiments were instructed in Finnish and all participants were fluent in Finnish.

Apparatus

The experiments were conducted at the driving simulator laboratory at the University of Jyväskylä. The driving simulator can be described as medium-fidelity with the CKAS Mechatronics 2-DOF motion platform. The simulator consisted of longitudinally adjustable seat, Logitech G27 force-feedback steering wheel and pedals (Figure 1). During the experiments, automatic transmission was used. Three 40" LED screens (95.6 cm x 57.4 cm, resolution 1440 x 900 pixels per screen) were used to display the driving scene. A rear-view mirror, a head-up display speedometer and a RPM gauge were displayed on the middle screen. Both side screens had side mirrors. For the occlusion trial, both sides of the steering wheel were equipped with a lever that revealed the driving scene. Each press revealed the driving scene for 500 milliseconds as in the original occlusion method of Senders, Kristofferson, Levison, Dietrich and Ward [25]. Continuous pressing of the lever kept the driving scene continuously visible. The driving simulation software was provided by Eepsoft (<http://www.eepsoft.fi/>). Driving log data was saved at 10 Hz.



Figure 1: Experimental setup and the position of the tablet.

The predefined routes that were used during the trials simulated real Finnish suburban roads located at Martinlaakso, Vantaa. The roads were the same as used in the study of Kujala and Mäkelä [15]. The text entry methods and the in-car tasks were implemented based on Carrio application (Figure 2), an in-vehicle infotainment system (<https://carrioapp.com/>) running on 7" Lenovo TB3-730X tablet. In order to make a search with the keyboard, the user needed to tap the search field to activate the keyboard, type the search phrase and tap the magnifying glass key. To activate the handwriting method (developed by <http://www.myscript.com/>), the user had to tap the handwriting icon, enter the letters one letter at a time and finally, tap the check mark icon. The handwriting method gave audio feedback, that is, repeated the written letter out loud, and thus, enabled writing without visual attention. The voice recognition search was activated by tapping the microphone icon. For all the methods, the system listed several search results to choose from by tapping the result. Ergoneers' Dikablis 50 Hz head-mounted eye-tracking system was used to record participants' eye movements and a LAN bridge was used for the synchronization of the driving simulator (x, y, speed) and the eye-tracking data.

Procedure

The demographic data was collected before the experimentation via email. The participants signed an informed consent form before participating. Before the actual experiment, participants practiced driving in an artificial city environment with other road users. They were instructed to drive as long as they wanted, with an average practice time of 3.0 minutes. After they felt comfortable with driving, they started practicing for the occlusion drive: how to drive vision occasionally occluded and how to use the levers that removed the occlusion and revealed the driving scene. The practice took place in a same city environment with other road users as the previous practice. Mean practice time was 3.65 minutes.

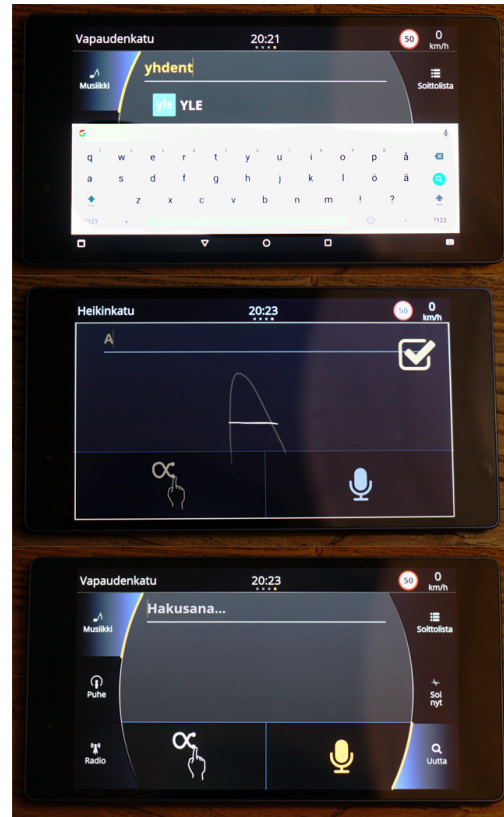


Figure 2: Three different text entry user interfaces: keyboard (on top), handwriting and voice recognition.

After the practices, the experiment started with the occlusion trial for test sample validation. In the trial, the screens were blank by default and the participants were able to see the driving scene for 500 milliseconds by pressing the levers on the steering wheel. In the trial, the participants were instructed to follow the traffic rules, to drive safely and at the same time to drive without visual information (i.e. vision occluded) as long as possible. An extra movie ticket was promised to those six drivers who could drive the longest periods without visual information but still accurately. This was done in order to make the participants to focus on the driving task but still trying to maximize the occlusion distance to their preference. A highway route without other road users was used in the occlusion trial. The route was the same as for the baseline sample of $N = 97$ in Kujala et al. [16]. The speed limits varied from 60 to 80 to 120 kilometers per hour during the trial and each change in a limit was given at the same point of the route by the experimenter. After the trial, each participant filled out the NASA-TLX questionnaire [7].

After the occlusion trial, the Visual Patterns Test [4] was completed. Once the test was done, the eye-tracking headset was put on, adjusted and calibrated and then the distraction test part started. In the distraction testing, the participants were instructed to prioritize the driving task, to obey the traffic regulations and to drive safely. The speed limit was

set to 50 kilometers per hour, but the participants were able to adjust the speed freely if needed.

Each participant completed three different tasks with three different text entry methods: keyboard, handwriting and voice recognition. The tasks were:

1. To write and find three different addresses
2. To write and start to play three different songs
3. To write and find three different contact information.

All tasks were completed with each text entry method. The order of the tasks and the driven routes were counterbalanced in order to avoid learning effect. The visual demands of the used routes were as similar as possible (as measured by OD in [15]) and there were no other road users on the routes. The traffic lights were always green when participant approached to junctions. After each text entry method tasks, the NASA-TLX questionnaire was filled out. Every participant was rewarded with a movie ticket after the experiment.

Analyses

The main dependent variables in the distraction testing were the percentage of red in-car glances (in-car glance distances above 85th percentile ODs for the 1x1-meter route points), the percentage of green in-car glances (in-car glance distances below or at median ODs for the 1x1-meter route points) and reduced NASA-TLX (no weighting) for each text entry method. In addition, we compared the number of the in-car glances and the number of errors (i.e. incorrectly recognized input and typing errors for keyboard) per text entry method. Drivers' occlusion distances (m) and in-car glance distances (m, distance traveled during an in-car glance) were measured for sample validation.

The in-car glance lengths were scored in real-time by a script that read the pupil's x and y coordinates from the eye-tracker. The coordinates were synchronized with the location data that the driving simulator provided. The glance lengths were scored following the SAE-J2396 [26] definition, with the exception that the gaze transition time back to the driving scene was added to a glance, in order to enable more direct comparability with the occlusion distance (no focal visual information available from the road during an in-car glance). All glances were manually searched from a synchronized video (25 fps) for validity using Noldus Observer XT software. All inaccuracies were manually corrected frame-by-frame.

The verification threshold for the red glances was set to 6 % and to 68 % for the green glances, based on Kujala et al. [14]. The percentage thresholds are based on the median percentages of the occlusion distances below or at the median OD ('green occlusions') and the occlusion distances exceeding the 85th percentile OD ('red occlusions') of the 97 drivers in Kujala et al. [16]. To test the equality of the median red and green in-car glance percentages of the three text entry methods to the verification thresholds (6 % and 68 %), one-sample sign test was used due to the highly non-

Gaussian distributions. The differences between the text entry methods were analyzed with Wilcoxon signed-rank test. In order to test the differences in the experienced task workload between the three text entry method trials and the occlusion trial, one-way repeated measures ANOVA was used. When the sphericity assumption was violated, the Greenhouse-Geisser correction was applied. Bonferroni corrections were applied for pairwise comparisons. Cronbach's Alpha was used to test the correlation and covariance between in-car glance distances across the different text entry tasks.

In the occlusion trial, the dependent variable was occlusion distance (OD). Because the occlusion metrics were non-Gaussian, median was used instead of mean. In order to control the effects of accelerations and decelerations in the beginning, in the intersections and in the end of the trial, only occlusion distances that were driven over 20 m/s (72 km/h) were included in the data. The Pearson product-moment correlation was used to test the correlation between median occlusion distance and median in-car glance distance as well as the correlation between median occlusion distance and the Visual Patterns Test scores [4]. The equality of the test drivers' OD distribution to the OD distribution of $N = 97$ in Kujala et al. [16] was assessed by Levene's test of equality of variances.

EXPERIMENT 1 - RESULTS

Number of glances and errors

Mean number of in-car glances during the nine tasks per method was 120 ($SD = 31$) for the keyboard, 201 ($SD = 61$) for the handwriting, and 84 ($SD = 26$) for the voice recognition. All the differences were significant ($p < .001$). There was a significant difference between the number of errors in the keyboard ($M = .9$, $SD = 1.3$) and the handwriting ($M = 4.2$, $SD = 2.5$) tasks ($Z = 3.304$, $p = .001$) as well as between the handwriting and the voice recognition ($M = 1.2$, $SD = 1.1$) tasks ($Z = 3.225$, $p = .001$). No difference was found between the keyboard and the voice recognition tasks ($p = .298$).

Red in-car glances

The verification threshold of the red glances was set to 6 % (at or below). The keyboard tasks did not pass the verification criteria, one-sample sign test indicating that the percentage of red glances was significantly higher than 6 % ($p = .003$, median = 13.22 % (Figure 3)). Either did the handwriting tasks, the percentage of red glances being also significantly higher than 6 % ($p = .001$, median = 9.49 %). The voice recognition tasks passed the verification criteria, the median percentage being 3.51 % ($p = .722$). Wilcoxon signed-rank test indicated that there was no difference in the percentages of red glances between the keyboard and the handwriting tasks ($Z = 1.349$, $p = .177$). However, there was a significant difference between the keyboard and the voice recognition tasks ($Z = 3.337$, $p = .001$) as well as the handwriting and the voice recognition tasks ($Z = 2.864$, $p = .004$).

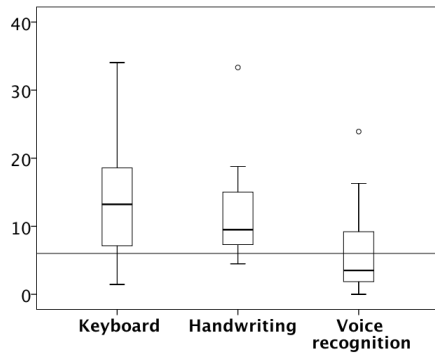


Figure 3: Percentage of red in-car glances per text entry method (the verification threshold at 6 %).

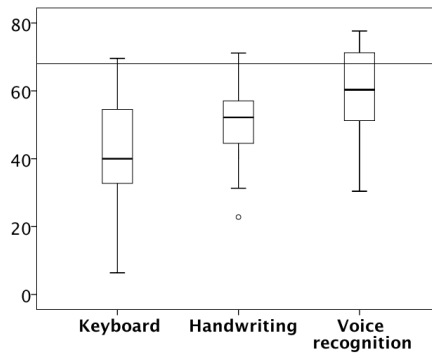


Figure 4: Percentage of green in-car glances per text entry method (the verification threshold at 68 %).

Green in-car glances

The verification threshold of green glances was set to 68 % (at or above). Only the voice recognition tasks passed the verification criteria, by not differing significantly from 68 % (Figure 4). The median of green in-car glances in the voice recognition tasks was 60.35 % ($p = .055$) whereas in the keyboard tasks median was 40.00 % ($p < .001$) and in the handwriting tasks 52.20 % ($p < .001$). After Bonferroni correction, there was no significant difference in the percentages of green glances between the keyboard and the handwriting tasks ($Z = 1.965, p = .049, \alpha = .017$). However, there was a significant difference between the keyboard and the voice recognition tasks ($Z = 3.432, p = .001$) as well as the handwriting and the voice recognition tasks ($Z = 2.580, p = .010$).

Experienced task workload - NASA-TLX

A significant main effect of trial was found, $F(2.174, 38.783) = 12.819, p < .001, \text{partial } \eta^2 = .445$. The handwriting tasks were experienced more demanding than the keyboard (mean difference 15.44, $p < .001$) and voice recognition tasks (mean difference 22.55, $p < .001$, Figure 5). After Bonferroni correction, the difference between the handwriting tasks and the occlusion trial was not significant ($p = .031, \alpha = .0083$). The difference between the keyboard tasks and the occlusion trial was also not significant ($p = .104$). The voice recognition tasks were experienced significantly less demanding than the occlusion trial ($p = .006$).

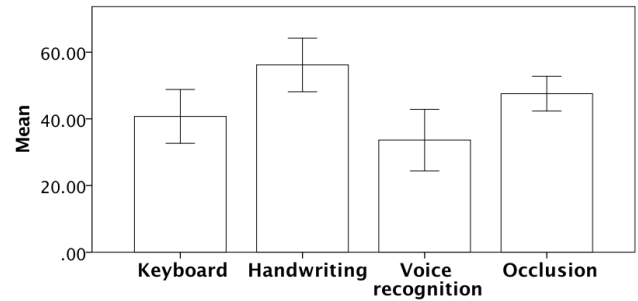


Figure 5: Experienced workload (NASA-TLX, max 100). Error bars: 95% CI.

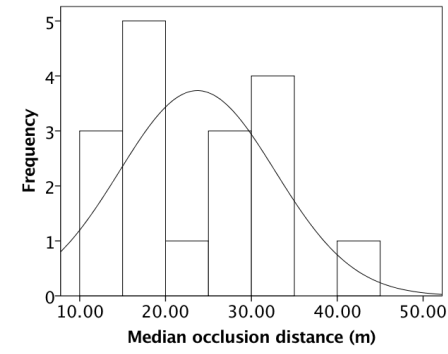


Figure 6: Individual median occlusion distances (m) on highway (speed > 72 km/h)

Occlusion distances

For sample validation, the distributions of the drivers' median ODs (Figure 6) were compared to the median ODs of the baseline data ($N = 97$; [16]). Drivers' median OD varied from 11.3 to 43.0 meters (median of 21.5 m). Levene's test indicated that the variance of the OD distribution does not differ significantly from the baseline OD distribution of $N = 97$ in Kujala et al. [16] ($F = 1.07, p = .303$) with a range between 3.2 to 41.9 meters.

In-car glance distances across the tasks, occlusion distance and visual short-term working memory

With each text entry method, three different types of tasks (3x3 tasks) were conducted: entering an address, entering a song, and entering a contact entry. High correlations between the 9 tasks were found and the Cronbach's alpha was excellent ($\alpha = .901$). However, there was no correlation between the occlusion distances and the in-car glance distances ($r = -.002, p = .994$). No correlation was found either between median occlusion distances and Visual Patterns Test scores ($N = 20, r = .232, p = .324, [4]$).

EXPERIMENT 1 - DISCUSSION

The visual distraction potential of three different text entry methods was studied following the testing and verification criteria of Kujala and Mäkelä [15]. Only the voice recognition-based text entry tasks passed the set verification criteria. The percentage of red in-car glances during the voice recognition tasks (3.51 %) was significantly lower than the verification threshold of 6 % as well as that of the keyboard (13.22 %) or handwriting (9.49 %) tasks.

Previous studies have shown similar results concerning the differences between voice recognition and touch screen keyboards (e.g., [3,9,23,27,28]).

The experienced task workload was the highest for the handwriting tasks, even higher than for the occlusion trial. The novelty of the handwriting method as well as the number of recognition errors by the system could explain some of the experienced workload. However, the method shows some promise as the percentage of red glances stayed at the same or even lower level than that for the keyboard tasks, even if there were significantly more errors (and thus, more glances) for the former. With higher recognition accuracy and more experienced users the method could be visually significantly less demanding than keyboard text entry. The voice recognition tasks were experienced as least demanding of all the tasks. During these tasks, manual input was considerably less needed than during the keyboard or the handwriting tasks.

The distribution of the occlusion distances (OD) was fairly similar to the baseline data [16] and the differences in the individual OD preferences can be assumed to not have affected the results of the distraction testing. Drivers with low ODs are not over-presented in the sample compared to the baseline data. The results of the distraction testing are well in line with earlier test results on significantly less demanding audio-visual route guidance (0.0-2.5 % red glances, Kujala et al., [14]).

The individual in-car glance length preference was found across the different in-car tasks as in previous studies [1,2,14,18,21]. However, correlation between the occlusion distances and the in-car glance distances was not found. In Kujala et al. [14] a correlation was found, but the locations of the in-car glances were more controlled (to follow route guidance) and the sample size was $N = 24$, here $N = 17$. In future studies, a more accurate metric to study this association would be, for instance, the ratio between in-car glance distance and the median OD of the baseline data ($N = 97$) on the route point where the glance is started. This would control the variability of the visual demands on the route points where the in-car glances are initiated. No correlation between OD and short-term visual memory capacity was found. Again, the sample size ($N = 20$) was small but it is unlikely there is more than a weak association between these two measures. More research is needed in order to explain the individual OD and in-car glance distance preferences.

In order to test our hypothesis on the effects of experience on eyes-on-road text entry with the handwriting method, we conducted another experiment focusing on this method. In addition, we wanted to study further the relationship between ODs and in-car glance lengths in text entry tasks.

EXPERIMENT 2 - INSTRUCTED HANDWRITING

In Experiment 2 we hypothesized that the handwriting method would have significantly lower visual distraction

potential if the drivers would have practiced the use of the method without vision. Again, the handwriting method gave audio feedback, that is, repeated the written letter out loud after each entry. We studied the question with 24 new participants, and compared the test results to those of Experiment 1.

Participants

The NHTSA [20] recommendations on the driver sample were followed as accurately as possible. The recruitment of the participants was done via university's mailing lists. In total 24 participants took part in the experiment: 17 males and 7 females. Five women indicated symptoms of simulator sickness and were replaced with male participants.

The age of the participants ranged from 20 to 79 years, mean age being 34.8 years ($SD = 16.0$). Eight of the participants were 18 to 24 years old, nine 25 to 39 years old, four 40 to 54 years old and three were older than 55 years. All participants had a valid driver's license and drove at least 5 000 kilometers per year. The total kilometers driven per year varied from 5 000 to 30 000, with a mean of 12 938 kilometers ($SD = 7 046$) per year. Their driving experience varied from 2 to 55 years, with a mean of 16.0 ($SD = 15.0$) years. Normal or corrected-to-normal vision was a prerequisite for participating. The experiments were instructed in Finnish and all participants understood and spoke Finnish. The participants were rewarded with a gift certificate (15 EUR) for participating the study.

Apparatus

The experiments were conducted at the driving simulator laboratory at the University of Jyväskylä and the used apparatus was the same as in Experiment 1. The used routes during the trials simulated real Finnish highways located at Martinlaakso, Vantaa and were the same as the ones used in the study of Kujala et al. [16]. This time, highway routes were used in order to keep the environmental visual demands of the driving as static and similar as possible for the analysis of the association between ODs and in-car glance distances. During the trials, no other road users were on the routes.

Procedure

The demographic data was collected in advance via email. An informed consent form was signed before the experiment. The practices were conducted similarly as in Experiment 1. The mean driving practice time was 5.79 minutes and the mean occlusion trial practice time was 4.33 minutes. The experiment started with the occlusion trial for test sample validation. The instructions were exactly the same as in Experiment 1. After the occlusion trial, NASA-TLX questionnaire [7] was filled out, the eye-tracking headset was put on, adjusted and calibrated and the distraction testing for the handwriting task started.

The participants were shown how the handwriting method is applied and how to write without glancing at the tablet's

screen (see Figures 1 and 2 [middle]). After the demonstration was their turn to repeat the exercise and rehearse to write without glancing at the screen (simulator stationary). The experimental task was to write an address using the handwriting method. The participants received an additional instruction to try to avoid glancing at the screen while writing and driving. The nominal speed limit changed from 120 to 80 kilometers per hour in the middle of the route after changing the road via junction, and this change was told to each participant at the same point of the route. They were also advised that they can freely adjust the speed if necessary. The route was a highway route with no other road users and every participant drove the same distance from the starting point to the ending point. During the drive, they wrote as many address entries as they could but in practice, two turned out to be the maximum number of addresses that a participant was able to finish. After the trial, the NASA-TLX [7] questionnaire was filled out and they were rewarded with a gift certificate.

Analysis

The main dependent variables in the distraction testing were the same as in Experiment 1. In addition, visual demand ratio, the ratio between in-car glance distance and the median OD of the baseline data ($N = 97$, Kujala et al, [16]) on the route point where the in-car glance is started, was measured. All the statistical analyses were conducted in the same manner as in Experiment 1 (for a single condition).

EXPERIMENT 2 - RESULTS

Number of glances and errors

The mean number of glances during the handwriting task per participant was 44 ($SD = 23$) and the mean number of errors per participant was 3.5 ($SD = 1.9$).

Red in-car glances

Again, the handwriting task did not pass the set verification criteria for the red in-car glances (Figure 7). One-sample sign test indicated that the percentage of red glances was significantly higher than 6 % ($p = .036$, median = 9.00 %).

Green in-car glances

The handwriting task did not pass the verification criteria for the green in-car glances (Figure 8). The percentage of the green glances was significantly lower than 68 % ($p < .001$, median = 52.50 %).

Experienced task workload - NASA-TLX

Wilcoxon signed rank test indicated that the handwriting task ($M = 57.29$, $SD = 12.17$) was experienced significantly more demanding than the occlusion trial ($M = 49.93$, $SD = 12.92$, $Z = -2.173$, $p = .030$, $d = .587$).

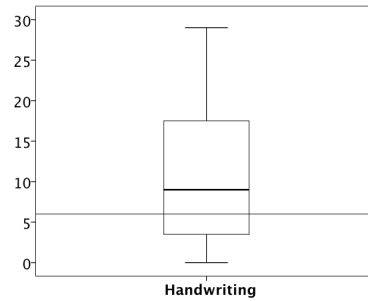


Figure 7: Percentage of red in-car glances (the verification threshold at 6 %).

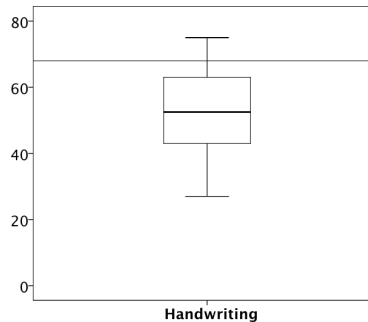


Figure 8: Percentage of green in-car glances (the verification threshold at 68 %).

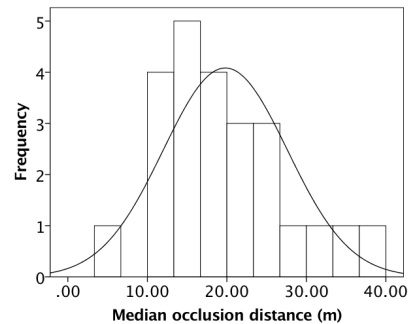


Figure 9: Individual median occlusion distances (m) on highway (speed > 72 km/h)

Occlusion distances

The median OD of the drivers varied from 6.35 to 37.18 meters, median being 17.98 meters (Figure 9). Levene's test indicated that the variance of the OD distribution does not differ significantly from the baseline OD distribution of $N = 97$ in Kujala et al. [16] ($F = .08$, $p = .778$). No correlation was found neither between OD and in-car glance distance ($r = -.193$, $p = .366$) nor between OD and visual demand ratio ($r = -.284$, $p = .179$). A strong correlation between in-car glance distance and visual demand ratio was found, $r = .824$, $p < .001$.

EXPERIMENT 2 - DISCUSSION

The visual distraction potential of the handwriting text entry method was re-evaluated with 24 drivers getting practice and instructions on eyes-on-road writing. Surprisingly, the results were highly similar to the findings in the first experiment. In the Experiment 1, the percentage of the red

glances was 9.49 % and in Experiment 2, the percentage was 9.00 %. The percentage of the green glances in the first experiment was 52.20 % and in the second experiment 52.50 %. The percentages indicate that the handwriting as a text entry method did not pass the set verification criteria in either experiment. It seems that the rather short practice of writing without glancing at the screen did not work that effectively for minimizing the tasks' visual distraction potential. Perhaps longer experience in using the handwriting method could improve the skill to write without watching the screen while driving. The relatively high number of character recognition errors could also have affected the percentage of the red in-car glances (Experiment 1: mean 4.2, Experiment 2: mean 3.5). The number of errors during the handwriting task was still quite high in Experiment 2 despite of the practice to write eyes-on-road.

As previously, the experienced task workload was higher during the handwriting task than during the occlusion trial. Again, we assume that the high number of recognition errors led to high levels of experienced task workload due to higher visual demand of the task in the form of additional glances for making corrections. Predictive text input [19,29], allowing for more inaccurate input for individual characters, could significantly decrease the visual distraction potential of the method but this should be further studied. A limitation of the used testing environment is the degrees of freedom in the movements of the motion platform (2 DOF). Road surface roughness and other vertical movements of the vehicle, which could further affect the usefulness of the handwriting method (as well as touch screen keyboard) on real roads, were absent in the driving simulation.

The distribution of the occlusion distances (OD) was again fairly similar in Experiment 2 as in the baseline data of Kujala et al. [16], probably due to the inclusion of different age groups (older drivers preferring shorter ODs, $r = -.437$, $p = .037$, $N = 23$). It can yet again be assumed that the individual OD preferences did not affect the results of the distraction testing. Most importantly, low OD drivers were not over-represented in the sample compared to the baseline data or the sample in Experiment 1.

The effects of the varying visual demands were better controlled in Experiment 2 than in Experiment 1 due to the highway routes. This is evident from the strong correlation between the in-car glance distances and the visual demand ratio, indicating that the visual demands did not vary significantly between the in-car glances. Yet again, OD did not correlate with in-car glance lengths. We assume that the missing correlation is due to variability in the participants' capabilities in the writing task and the nominal speed limits. The low-OD drivers (aged, in particular) may have required longer in-car glances than the more skilled writers but they were not able to compensate sufficiently for this by decreasing driving speed due to the instructed speed limits

[6]. Secondary task related skills and also the structural constraints (e.g., natural break points, [8]) of an in-car task might affect the in-car glance lengths more than the individual uncertainty that may rarely rise to the same level with these kinds of in-car tasks than in an occlusion experiment (cf. the percentages of green in-car glances). An open question remains, if the sample should be validated based on the OD or the in-car glance length distributions. As the metric of red in-car glances is based on ODs and it seems to provide reliable results, we suggest the former is more important for this type of distraction testing.

The highly similar distraction test results for the handwriting tasks between Experiment 1 and Experiment 2 provide reliability for the suggested testing method and verification criteria. The test results were highly similar even if the participant sample, the road environment (suburban vs. highway), and driving speeds were different. This suggests that comparable test data could be gathered with the testing method even if it is applied to different driving simulator and driving scenario implementations, in which the baseline occlusion data can be collected.

CONCLUSIONS

The visual distraction potential of three different text entry methods was studied following the testing and verification criteria of Kujala and Mäkelä [15]. Only the voice recognition-based text entry tasks passed the set verification criteria based on the percentages of red and green in-car glances (in-car glance lengths above 85th percentile or below median of the baseline ODs, correspondingly). The percentage of red in-car glances during the voice recognition tasks (3.51 %) was significantly lower than the verification threshold of 6 % as well as that of the keyboard (13.22 %) or handwriting (9.49 %) tasks. The voice recognition tasks were also experienced as least demanding of all the tasks.

The handwriting method was further studied with 24 participants with instructions and practice on writing with eyes on road. The practice on the method did not affect the test results significantly. The findings suggest that handwriting could be visually less demanding than touch screen typing but the reliability of the text input recognition should be significantly improved or the driver well-experienced with the method in order to minimize its visual distraction potential. The handwriting method could be further researched with participants who are already familiar with using the method.

The highly similar distraction test results for the handwriting tasks between Experiment 1 and Experiment 2 provide reliability for the suggested testing method and verification criteria.

ACKNOWLEDGEMENTS

This work was funded by Ficonic Solutions Ltd. The funder had no role in data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1. Robert Broström, Peter Bengtsson and Mikael Ljung Aust. 2016. Individual glance strategies and their effect on the NHTSA visual manual distraction test. *Transportation research part F: traffic psychology and behaviour*, 36, 83-91. <http://dx.doi.org/10.1016/j.trf.2015.10.017>
2. Robert Broström, Mikael Ljung Aust, Linnea Wahlberg and Laban Källgren. 2013. What drives off-road glance durations during multitasking—capacity, practice or strategy. In *Proceedings of the 3rd International Conference on Driver Distraction and Inattention*.
3. Joseph M. Crandall and Alex Chaparro. 2012. Driver distraction: Effects of text entry methods on driving performance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 56,1: 1693-1697. <https://doi.org/10.1177/1071181312561339>
4. Sergio Della Sala, Colin Gray, Alan Baddeley, Nadia Allamano and Lindsey Wilson. 1999. Pattern span: a tool for unwelding visuo-spatial memory. *Neuropsychologia*, 37, 10: 1189-1199. [http://dx.doi.org/10.1016/S0028-3932\(98\)00159-6](http://dx.doi.org/10.1016/S0028-3932(98)00159-6)
5. James P. Foley, Richard Young, Linda Angell and Joshua E. Domeyer. 2013. Towards operationalizing driver distraction. In *Proceedings of the 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, 57-63.
6. Ray Fuller. 2005. Towards a general theory of driver behaviour. *Accident Analysis & Prevention* 37, 3: 461-472. <http://dx.doi.org/10.1016/j.aap.2004.11.003>
7. Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, 52, 139-183. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
8. Christian P. Janssen, Duncan P. Brumby and Rae Garnett. 2012. Natural break points: The influence of priorities and cognitive and motor cues on dual-task interleaving. *Journal of Cognitive Engineering & Decision Making*, 6, 5-29. <https://doi.org/10.1177/1555343411432339>
9. Jibo He, Alex Chaparro, Bobby Nguyen, Rondell J. Burge, Joseph Crandall, Barbara Chaparro, Rui Ni and Shi Cao. 2014. Texting while driving: Is speech-based text entry less risky than handheld text entry? *Accident Analysis & Prevention*, 72, 287-295. <http://dx.doi.org/10.1016/j.aap.2014.07.014>
10. Jibo He, William Choi, Jason S. McCarley, Barbara S. Chaparro and Chun Wang. 2015. Texting while driving using Google Glass™: Promising but not distraction-free. *Accident Analysis & Prevention*, 81, 218-229. <http://doi.org/10.1016/j.aap.2015.03.033>
11. Petri Jääskeläinen and Leena Pöysti. 2014. *Tarkkaamattomuus tieliikenteen turvallisuusongelmana – suomalaisten käsityksiä* (Driver distraction and inattention as a safety issue in road traffic – Finnish views). Helsinki: Finnish Road Safety Council.
12. Dagmar Kern, Albrecht Schmidt, Jonas Arnsmann, Thorsten Appellmann, Nillakshi Pararasasegaran and Benjamin Piepiera. 2009. Writing to your car: handwritten text input while driving. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*, 4705-4710. ACM. 10.1145/1520340.1520724
13. Katja Kircher and Christer Ahlstrom. (in press). Minimum required attention: a human-centered approach to driver inattention. *Human factors*. <https://doi.org/10.1177/0018720816672756>
14. Tuomo Kujala, Hilka Grahn, Jakke Mäkelä and Annegret Lasch. 2016a. On the Visual Distraction Effects of Audio-Visual Route Guidance. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 169-176. ACM. 10.1145/3003715.3005421
15. Tuomo Kujala and Jakke Mäkelä. 2015. Development of a testing environment and a verification procedure for in-car tasks with dynamic driving scenarios. In *Proceedings of the 4th International Conference on Driver Distraction and Inattention*.
16. Tuomo Kujala, Jakke Mäkelä, Ilkka Kotilainen and Timo Tokkonen, 2016b. The Attentional Demand of Automobile Driving Revisited: Occlusion Distance as a Function of Task-Relevant Event Density in Realistic Driving Scenarios. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 58, 1: 163-180. <https://doi.org/10.1177/0018720815595901>
17. Tuomo Kujala, Johanna Silvennoinen and Annegret Lasch. 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*, 82-89. ACM. 10.1145/2516540.2516562
18. Mikael Ljung Aust, Sergejs Dombrovskis, Jordanka Kovaceva and Bo Svanberg. 2013. An empirically based suggestion for reformulating the glance duration criteria in NHTSA's visual-manual interaction guidelines. *SAE International Journal of Passenger Cars-Electronic and Electrical Systems*, 6(2013-01-0444), 444-453. 10.4271/2013-01-0444
19. I. Scott MacKenzie and R. William Soukoreff. 2002. Text entry for mobile computing: Models and methods, theory and practice. *Human-Computer*

- Interaction*, 17, 2-3: 147-198.
10.1080/07370024.2002.9667313
20. National Highway Traffic Safety Administration. (2013). *Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices*. (NHTSA-2010-0053.) Washington DC, NHTSA.
 21. Jami Pekkanen, Otto Lappi, Teemu H. Itkonen and Heikki Summala. 2017. Task-difficulty homeostasis in car following models: experimental validation using self-paced visual occlusion. *PLoS one*, 12, 1: e0169704.
<http://dx.doi.org/10.1371/journal.pone.0169704>
 22. Bryan Reimer and Bruce Mehler. 2013. The effects of a production level “voice-command” interface on driver behavior: summary findings on reported workload, physiology, visual attention, and driving performance. MIT AgeLab Technical Report No. 2013-17A. Massachusetts Institute of Technology, Cambridge, MA.
 23. Bryan Reimer, Bruce Mehler, Ian Reagan, David Kidd and Jonathan Dobres. 2016. Multi-modal demands of a smartphone used to place calls and enter addresses during highway driving relative to two embedded systems. *Ergonomics*, 59, 12: 1565-1585.
<http://dx.doi.org/10.1080/00140139.2016.1154189>
 24. Aubrey Samost, David Perlman, August G. Domel, Bryan Reimer, Bruce Mehler, Alea Mehler, Jonathan Doblere and Thomas McWilliams. 2015. Comparing the Relative Impact of Smartwatch and Smartphone Use While Driving on Workload, Attention, and Driving Performance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 59, 1: 1602-1606. Sage CA: Los Angeles, CA: SAGE Publications.
<https://doi.org/10.1177/1541931215591347>
 25. John W. Senders, A. B. Kristofferson, W. H. Levison, C. W. Dietrich and J. L. Ward. 1967. The attentional demand of automobile driving. *Highway research record*, 195.
 26. Society of Automotive Engineers. 2000. *SAE-J2396 Definitions and Experimental Measures Related to the Specification of Driver Visual Behavior Using Video Based Techniques*. Warrendale, PA: SAE.
 27. Kathryn G. Tippey, Elayaraj Sivaraj, Wil-Johneen Ardoin, Trey Roady and Thomas K. Ferris. 2014. Texting while driving using Google Glass: Investigating the combined effect of heads-up display and hands-free input on driving safety and performance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58, 1: 2023-2027. Sage CA: Los Angeles, CA: SAGE Publications.
<https://doi.org/10.1177/1541931214581422>
 28. Omer Tsimhoni, Daniel Smith and Paul Green. 2004. Address entry while driving: Speech recognition versus a touch screen keyboard. *Human Factors*, 46, 4: 600-610.
<https://doi.org/10.1518/hfes.46.4.600.56813>
 29. Gabriela Villalobos-Zúñiga, Tuomo Kujala and Antti Oulasvirta. 2016. T9+ HUD: Physical Keypad and HUD can Improve Driving Performance while Typing and Driving. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 177-184. ACM. 10.1145/3003715.3005453