Assessing the effect of in-vehicle task interactions on attention management in safety-critical events

Bobbie Seppelt [1*], Sean Seaman [2], Linda Angell [2], Bruce Mehler [1], Bryan Reimer [1]

[1] AgeLab & New England University Transportation Center, Massachusetts Institute of Technology, 1 Amherst Street, Cambridge, USA
[2] Touchstone Evaluations, Inc., 440 Burroughs Street, Detroit, USA
[*]bseppelt@mit.edu

Abstract: Two analytic techniques were applied to study patterns of on- and off-road glances in naturalistic driving. The dataset used in this study was the Naturalistic Engagement in Secondary Task (NEST) database, a subset of the Strategic Highway Research Program (SHRP2) database, which contains safety-critical event (SCE) data comprised of Crash and Near-crash epochs curated so as to only contain incidents linked to secondary task activity. Output from an attention buffer, which produces a hybrid metric based on how on- and off-road glances are threaded over time, was analyzed in a comparison of safety-critical events to Baseline driving. Individual glance metrics of mean single glance duration (MSGD), number of glances, and proportion of glances by location, binned in 5-s intervals, were also analysed to diagnose the underlying behavioural patterns produced from the attention buffer. Statistical comparisons between SCEs and Baseline driving showed that regardless of secondary task type, during SCEs, drivers exhibited a destabilization of attention over time not evident in Baseline driving. Further examination of these effects based on an analysis of accumulated buffer loss revealed a more pronounced fracturing of attention over time for epochs containing visual-manual secondary task activity than those containing only auditory-vocal secondary task activity.

1. Introduction

A recent analysis of safety-critical events from the 100-Car Naturalistic Driving Study revealed the importance of on-road glance length in-between off-road glances in the moments preceding near-crash and crash outcomes [1]. In the 25s of time prior to these events, drivers involved in near-crashes (i.e., averted crashing) had significantly longer on-road glances, and looked less frequently between on- and off-road locations as compared to those involved in crashes. The authors showed that patterns of glance between on- and off-road locations differentiated safety-critical events (SCE) due to cumulative effects produced from the length of time drivers glanced to each location. These time-history effects were evident in consecutive time-bins of mean single glance duration (MSGD) and in output produced from the AttenD algorithm [2]. Based on these findings, the authors called for the use of metrics and analytic techniques that allow for a comparison of different glance sequences to multiple locations to complement existent assessment methods focused on single-region (commonly, off-road) glance allocation [3].

To further examine the extent to which the duration of on-road glances threaded between off-road glances produce patterns linked to safety-critical outcomes, the same analytic techniques introduced in [1] were applied to an analysis of a subset of SCEs from the Strategic Highway Research Program (SHRP2) naturalistic driving study [4] contained within The Naturalistic Engagement in Secondary Task database (NEST). The consideration of data from NEST allows for a more in-depth analysis on the extent to which the glance behaviours evident in the safety-critical epochs from the 100-car dataset are descriptive of a common pattern of attentional mismanagement in the moments prior to crashes and near-crashes, and/or, are preconditioned on interactions contingent on secondary task type. Unlike the 100-Car dataset, SCE epochs within NEST are all known to include secondary tasks. This additional coding of secondary activity enables an exploration of how task type disrupts glance behaviour in the moments prior to a precipitating event compared to Baseline driving. It is hypothesized that drivers engaged in secondary tasks display a destabilized glance pattern as compared to Baseline driving. Further, tasks that impose higher visual load are anticipated to produce increased destabilized patterns compared to those which primarily draw upon cognitive resources [5].

2. Method

The dataset used in this study was the Naturalistic Engagement in Secondary Task (NEST) database [4], a subset of the Strategic Highway Research Program (SHRP2) database, containing safety-critical event (SCE) data comprised of Crash and Near-crash epochs curated so as to only contain incidents linked to secondary task activity, as well as four Baseline epochs (i.e., epochs that do not contain SCEs) from each driver for each of that driver's independent observations in the SCE set. All the SCE epochs contain secondary task activity, which we categorized as visual-manual (e.g., any reaching, adjusting, manipulating, or holding activity), auditory-vocal (e.g., any conversation activity with a passenger, on the phone, or via voice commands to an in-cab system), or "mixed-mode," containing both kinds of secondary task activity (see Table 3 in Appendix A for a list of secondary tasks in NEST and how they were categorized, as well as how many epochs were observed for each type of SCE). Baseline epochs contained a mixture of those containing secondary task activity and those without, in order to reflect a truly random sampling of behaviour for those drivers found in the NEST SCE set. In the following analyses, "Baseline" values are always drawn from this mix of epochs, some of which contain secondary tasks, some of which do not. For example, when Crash epochs containing auditory-vocal tasks are compared to Baseline epochs, the comparisons are made within-subject, but behaviours observed are limited to those Crashes containing auditory-vocal tasks, while all Baselines are aggregated regardless of secondary task activity present, so as to compare behaviours during SCEs that are potentially linked to categories of secondary task behaviour to drivers' own typical behaviours (i.e., randomly selected) in routine driving.

Crash and Near-crash epochs were selected from exclusive groups of drivers, because, in NEST, Crash epochs outnumber Near-crash epochs. In cases where a single driver had both Crash and Near-crash epochs, the Crash epochs were removed, so that all statistics were computed on independent samples. This filtering yielded a set of 78 Near-crash epochs, 133 Crash epochs, and 940 Baseline epochs. For visualizations and statistical comparisons, epochs were further aggregated within drivers (because a single driver occasionally appeared in multiple SCEs of the same type, and always appeared in multiple Baseline epochs), yielding a set of 67 Near-crash drivers, 127 Crash drivers, and equivalent Baseline epochs.

For analyses utilizing the attention buffer, this set was further reduced by eliminating epochs that did not contain at least 19 seconds of glance data. The set was still further reduced by removing epochs from the SCE sets that did not have corresponding epochs in each driver's matched Baseline set; each secondary task grouping (Auditory-vocal, Visual-manual, and Mixed-mode) contained epochs from both SCE and Baseline sets for each driver in order to compute within-subject comparisons between Baseline and SCE. The dataset was further trimmed so that Crash and Near-crash epochs contained fully non-overlapping sets of drivers. This further filtering yielded a set of drivers, organized by task composition of epochs, shown in Table 1.

Table 1 Number of drivers, by SCE type and task composition for attention buffer analyses

	Near-crash	Crash
Auditory-vocal	5	17
Mixed-mode	12	29
Visual-manual	35	36

The primary behaviour of interest was glancing: In NEST, glance behaviour is provided in a sample-by-sample format, at 10 Hz, with each sample coded with an area-of-interest. For SCE epochs, only glance data prior to the onset of the precipitating event of the SCE was used, up to 20 seconds; for Baseline epochs, entire epochs were used, up to 20 seconds. Epochs that did not contain at least 19 seconds of data were excluded; thus, the entire data set consisted of 20 second epochs that either entirely preceded an SCE or was routine (Baseline) driving drawn from the sample of SCE drivers. From these periods of glance behaviour, four glance statistics were computed: mean single glance duration (MSGD), number of glances, proportion of glances to a location, and mean attention buffer value. Off-road locations in the vehicle that were designated as irrelevant for driving-related situation awareness included the driver's cell phone, iPod, or other interior objects, the centre stack, passengers, over-the-shoulder, or periods of time where the eyes were closed or were otherwise clearly off-road, even if not visible. Off-road locations in the vehicle that were designated as relevant to driving-related situation awareness included the instrument cluster, rear-view mirror, and left and right windows or side mirrors. On-road peripheral locations included the left and right windshield, while the main on-road location was coded as forward. For all three of the typical glance measures (MSGD, # of glances, and proportion of glances to a location), values were averaged first within drivers across available epochs, and then across drivers.

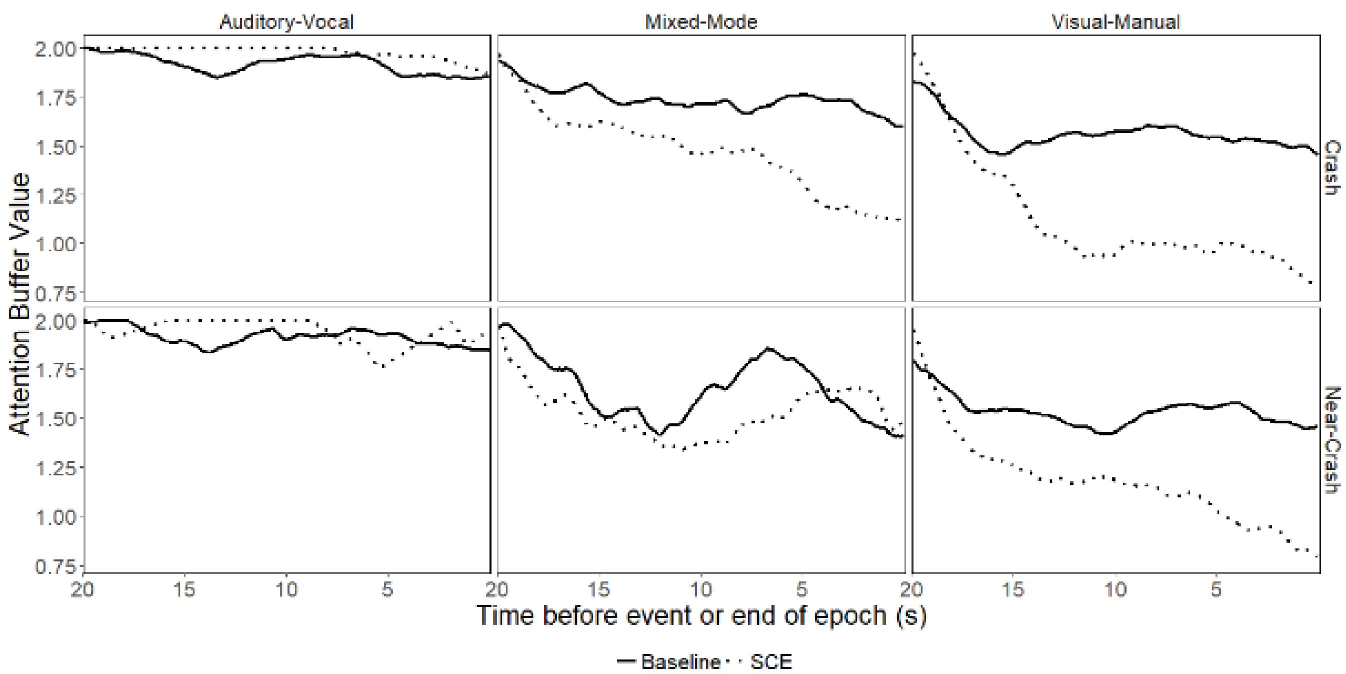

Fig. 1. Attention buffer by type of SCE and secondary tasks

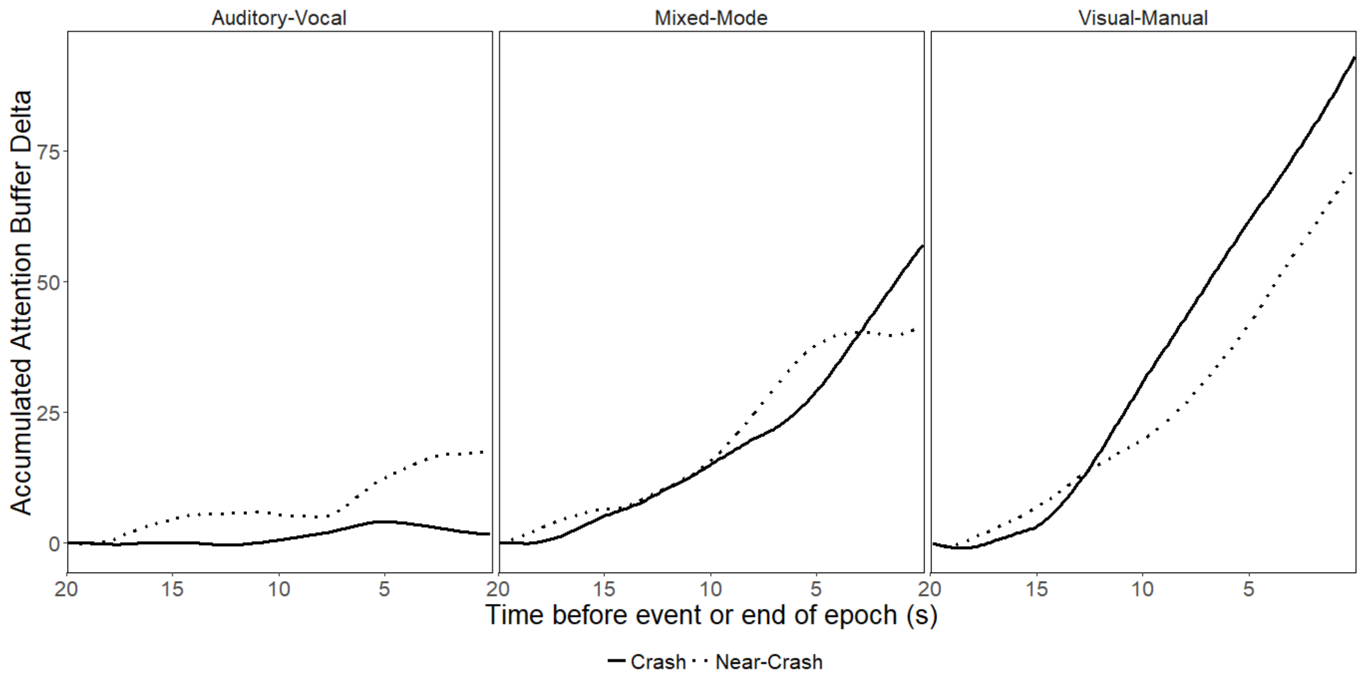

Fig. 2. *Accumulated difference in attention buffer between SCE and Baseline by SCE type and secondary task composition*

Averages were plotted with standard error of the mean bars to reflect the variance across drivers.

For the attention buffer measure, a modified form of the AttenD algorithm, first described within [6], was applied on an epoch-by-epoch basis. In its modified form, the Attention Buffer represents the amount of stored information about the roadway. Its value is tied to processes of attention and memory that are at play in how drivers sample information to form, retain, and update a robust representation of the driving environment [1]. At the start of each epoch, the initial buffer value was set at 2. For each second of off-road glance, the buffer value was decremented by 1 point. If the AttenD value reached 0, it did not drop further until the driver glanced back to the forward road, at which point it began increasing again, after a latency period of 0.2 seconds, reflecting an experimentally-derived minimum time required, following from an attentional shift, to perceive the presence and relative location of elements that have meaning for maintaining safe travel and anticipating potential hazards [7]. The rate of increment once glance returned to the forward road was set at a rate of 0.33 points per second, until it returned to 2 points. This rate specifies an average value corresponding to the amount of on-road glance time it takes to fully perceive and comprehend the presence of a slow-moving, non-salient, or peripherally-located hazard [8-12]. Glances to mirrors and the instrument cluster did not result in a decrement of the buffer until the duration exceeded 1 second, at which time the buffer decremented by 1 point per second. An up to 1-second time delay for these regions was included because they contribute to situationally-aware driving. Visualizations of the buffer data were made by averaging across epochs per type (i.e., near-crash, crash, baseline) for each time point within the 19-20 seconds (190-200 samples).

3. Results

Results are first presented for attention buffer analyses; later, differences between attention buffer profiles are explored in terms of traditional glance metrics.

Attention buffer scores were aggregated first by subject within each group of secondary tasks (Auditory-vocal, Visual-manual, and Mixed-mode), and then across drivers for each sample point in the 19-20 second period before a precipitating event (in SCE epochs) or the end of the epoch (in Baselines). Thus, each sample point becomes an average of averages, with more epochs aggregated in Baseline. Each SCE aggregated buffer line is plotted next to the aggregated Baseline buffer line from its matched drivers who had the same epoch secondary task composition within their Baseline periods. These plots can be seen in Fig. 1. Across the secondary task groupings, the slope of each buffer line, from the earliest moments before the end of an epoch, to the end of the epoch, tends to be negative, but changes in steepness as the task composition moves from Auditory-vocal, to Mixed-mode, to Visual-manual. For Auditory-vocal epochs, these lines, whether Near-crash or Crash, and whether Baseline or SCE, appear flat, suggesting there is no recorded loss of (visually-based) driving-related situation awareness across the span of the epoch. However, starting with Mixed-mode epochs, differences appear visible for Crash epochs between their SCE and Baseline counterparts, while less of a distinction appears for Near-crash epochs. For Near-crash Visual-manual epochs, the difference does appear, and the difference between SCE and Baseline attention buffer appears to be the greatest in magnitude between the Crash Visual-manual SCE and Baseline epochs.

Table 2 LME coefficients for attention buffer slope analyses

Sec. Task	Model Term	*B*	Std. Error	*t*
Visual-manual	Time	0.00320	0.00007	42.69***
	SCE Type	-0.07284	0.17570	-0.42
	Time x SCE Type	-0.00057	0.00015	-3.79***
Auditory-vocal	Time	-0.00001	0.00005	-0.12
	SCE Type	0.03491	0.05013	0.70
	Time x SCE Type	0.00023	0.00011	2.08*
Mixed-mode	Time	0.00161	0.00012	13.71***
	SCE Type	-0.15200	0.22660	-0.67
	Time x SCE Type	-0.00343	0.00026	-13.43***

* = $p < .05$; *** $p < .001$

To assess the statistical significance of these apparent differences in slope, linear mixed effects (LME) models [13] were computed, regressing the difference between drivers' aggregate Baseline buffer score and their SCE buffer score against the time point of each sample. These were computed separately, by task composition, and the interaction between time in epoch and type of SCE (Crash or Near-crash) was also assessed as a second-order effect. These results can be seen in Table 2 For each type of secondary task composition, the change in the attention buffer from matched Baseline driving, engaged in the same category of secondary tasks, displayed a significantly different slope over time as a function of whether that time period immediately preceded a Crash or a Near-crash. For Mixed-mode and Visual-manual epochs, this difference was due to a steeper slope in Crashes than Baseline, compared to Near-crashes and Baseline; for Auditory-vocal epochs, the effect was reversed, and far more subtle.

In addition to comparing the average difference, time point by time point, between SCE and baseline epochs, we also looked at the accumulation of this difference over time, in what can be interpreted as an area-under-the-curve, depicting the accumulated effect of aggregated loss of situation awareness versus Baseline driving within a secondary task modality. These effects are visualized in Fig. 2. Overall, the accumulated loss of (visually-mediated) driving-related situation awareness is greater in the Crash epochs containing Visual-manual tasks; this accumulated loss shows a steeper decline (shown here by a more positive slope) than Near-crash epochs of the same modality. LME analyses suggest that Auditory-vocal and Visual-manual accumulated attention buffer changes differ significantly over time between Crash and Near-crash epochs ($p < .001$ for both models).

These two sets of effects suggest that driver glance behaviour is different between Crash, Near-crash, and Baseline epochs, even when those epochs are controlled for both driver and the modality of secondary task composition. To better understand what specific glance behaviours may be driving these effects, we examined patterns in glances to different areas of interest across these groups using three measures: mean single glance duration, number of glances, and glance proportion.

For mean single glance duration, mean statistics were computed for on-road glances and off-road glances, as well as for Crash, Near-crash, and Baseline epochs; furthermore, statistics were computed separately for SCE epochs that contained Auditory-Vocal tasks, Visual-Manual tasks, or a mix of the two. Furthermore, glances were "binned" based on the time point at which the glance was initiated; for example, a glance initiated 18 seconds before the end of the epoch was placed in the 15-20 s bin. While long glances may straddle multiple 5 s bins, glances are only placed in the bin in which they are initialized; because glances can be long (especially on-road glances), mean glance duration tends to drop as bins get closer to the end of an epoch, due to the temporal limit on how long they can be sustained given the available window. Average glance duration for forward glances is presented in Fig. 3, and MSGD for other locations is presented in Fig. 4. Note that for each "Baseline" mean single glance duration value, it is the same across all types of task composition (because it represents typical, non-SCE driving performance randomly sampled from SCE drivers, and is being contrasted with SCE glance behaviour linked to different categories of secondary tasks).

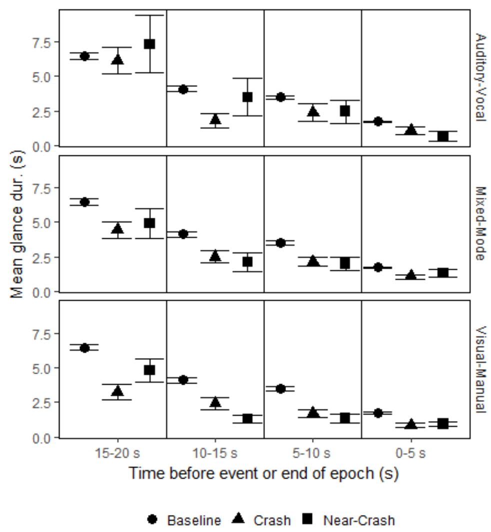

Fig. 3. *MSGD for forward glances by time to event, task modality, and SCE type. Error bars indicate standard error*

Average glance counts for each location are presented in Fig. 5, and average glance proportion—the proportion of each bin subtended by glances to a specific location—are presented in Fig. 6.

In comparing glance behaviour across Crash, Near-crash, and Baseline epochs, comparisons were done as repeated measures t-tests. Notably, *p* values were not

Bonferroni-corrected, as the available data within a cell was sparse and the number of comparisons was large; thus, the probability of a type I error is likely high. However, our goal was to examine the trends of glance differences within temporal bins, and to identify the bins with the greatest likelihood of being associated with significant differences in glance behaviour between SCE epochs and Baseline epochs. Thus, it is important to recognize that,

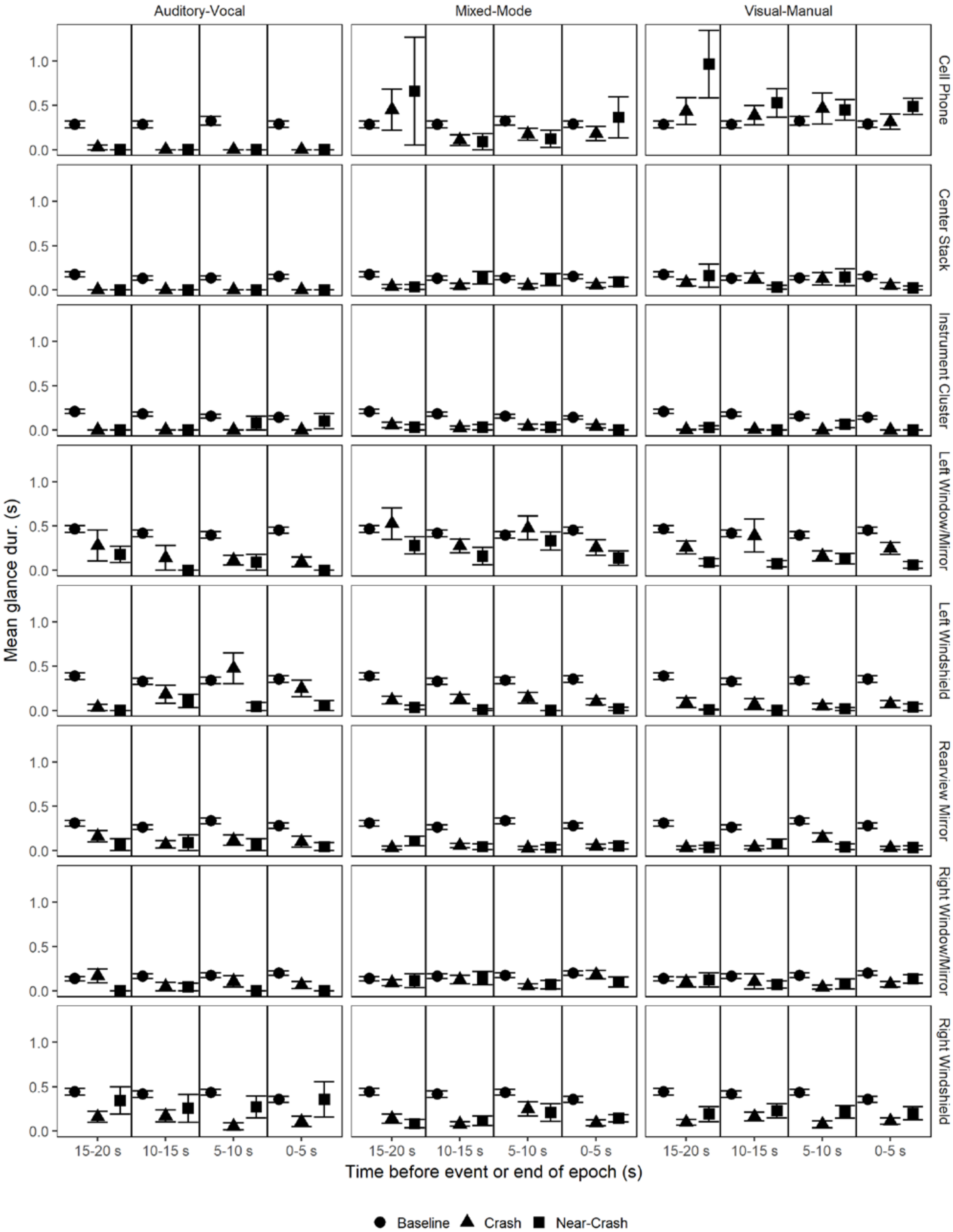

Fig. 4. *MSGD (s) by location, task modality, and SCE type. Error bars indicate standard error*

were the tests to be repeated on a new set of data, finding significant differences within any given bin with a similarly sized sample may not be successful; however, this binning approach provides a guide as to when differences emerge in the moments preceding precipitating events.

The greatest differences between SCE and Baseline glance duration occurred in the bins farthest away from the end of the epochs (i.e., farthest away from the precipitating event in SCE epochs): the 15-20 s bin, $t(33) = 2.35$, $p = .025$, and the 10-15 s bin, $t(36 = 2.75$, $p = .0093$. Smaller, but significant differences were observed in the 5-10 s bin, $t(34) = 2.15$, $p = .039$, and 0-5 s bin, $t(36) = 2.2$, $p = .034$. Near-crashes were associated with longer off-road glances in the 15-20 s bin $t(22) = 2.21$, $p = .038$, the 10-15 s bin $t(21) = 2.15$, $p = .044$, and the 5-10 s bin, $t(28) = 3.41$, $p = .002$). For Mixed-mode epochs, only the Crash 15-20 s bin, $t(35) = 1.78$, $p = .083$, and Crash 0-5 s bin, $t(39) =$

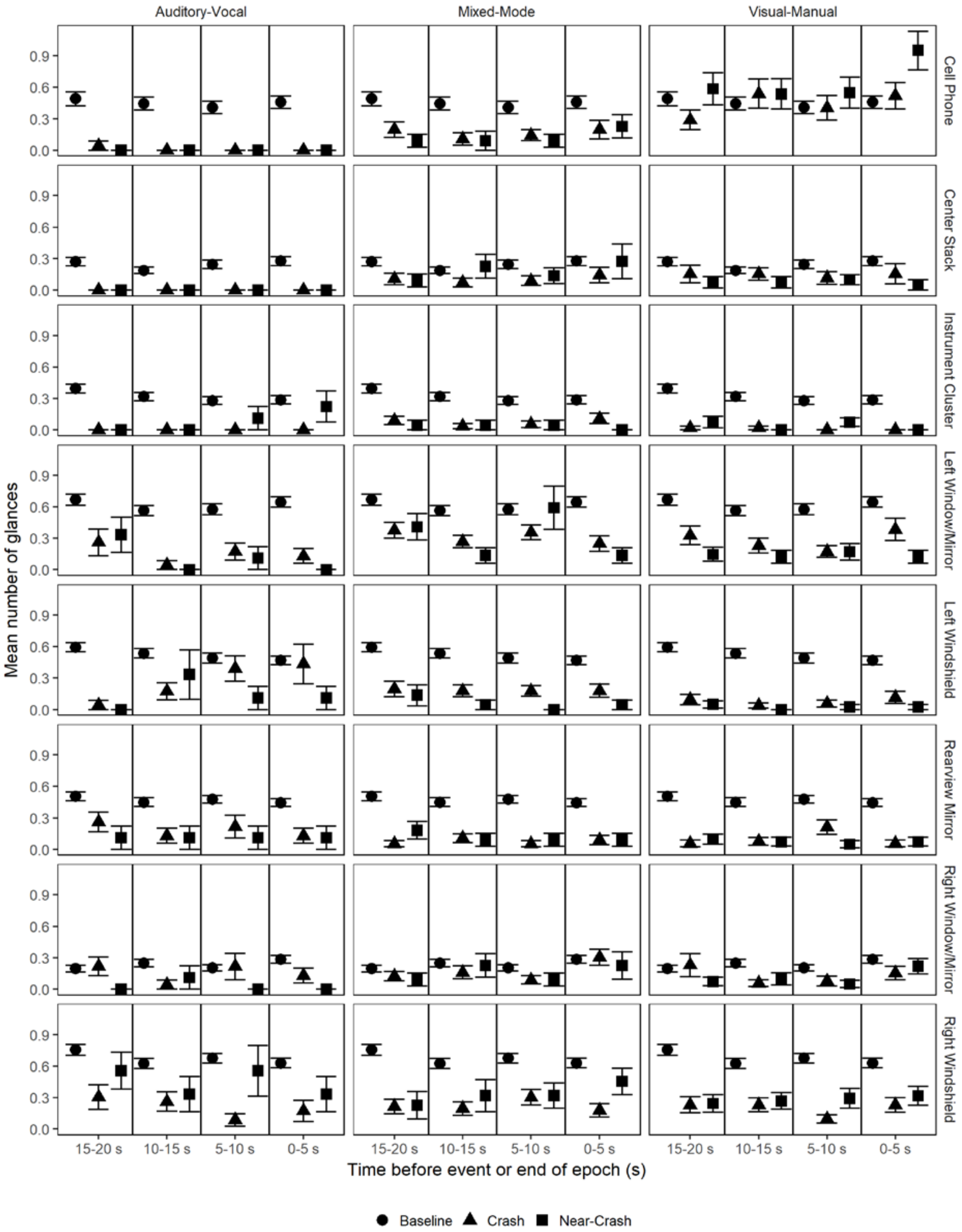

Fig. 5. Mean number of glances by time to event, location, task modality and SCE type. Error bars indicate standard error

1.73, p = .092, had marginally significant longer off-road glances than Baseline. No Near-crash off-road glances in any bin were significantly different than Baseline glances for Mixed-mode epochs. The only off-road difference observed in Auditory-vocal epochs were for Near-crashes, in the 5-10 s bin, $t(3)$ = 3.78, p =.03, with longer off-road glances being observed in baseline driving.

Mean on-road glances were shorter in Crash visual-manual than Baseline epochs in the 15-20 s bin, $t(48)$ = 2.12, p = .039, 5-10 s bin, $t(39)$ = 2.04, p = .049, and 0-5 s bin, $t(40)$ = 2.74, p = .0093; for Near-crash, significant differences were observed in the 5-10 s bin, $t(30)$ = 2.54, p = .017) and 0-5 s bin, $t(34)$ = 3.25, p =.0026, and a marginal difference was observed in the 10-15 s bin, $t(24)$ = 2.06, p = .051; notably there was no effect in the farthest bin, suggesting that one critical difference between Near-crash and Crash epochs containing visual-manual activity is that the differences in glance behaviour, compared with Baseline, extend only to time periods closer to the SCE. No significant differences were observed between Near-

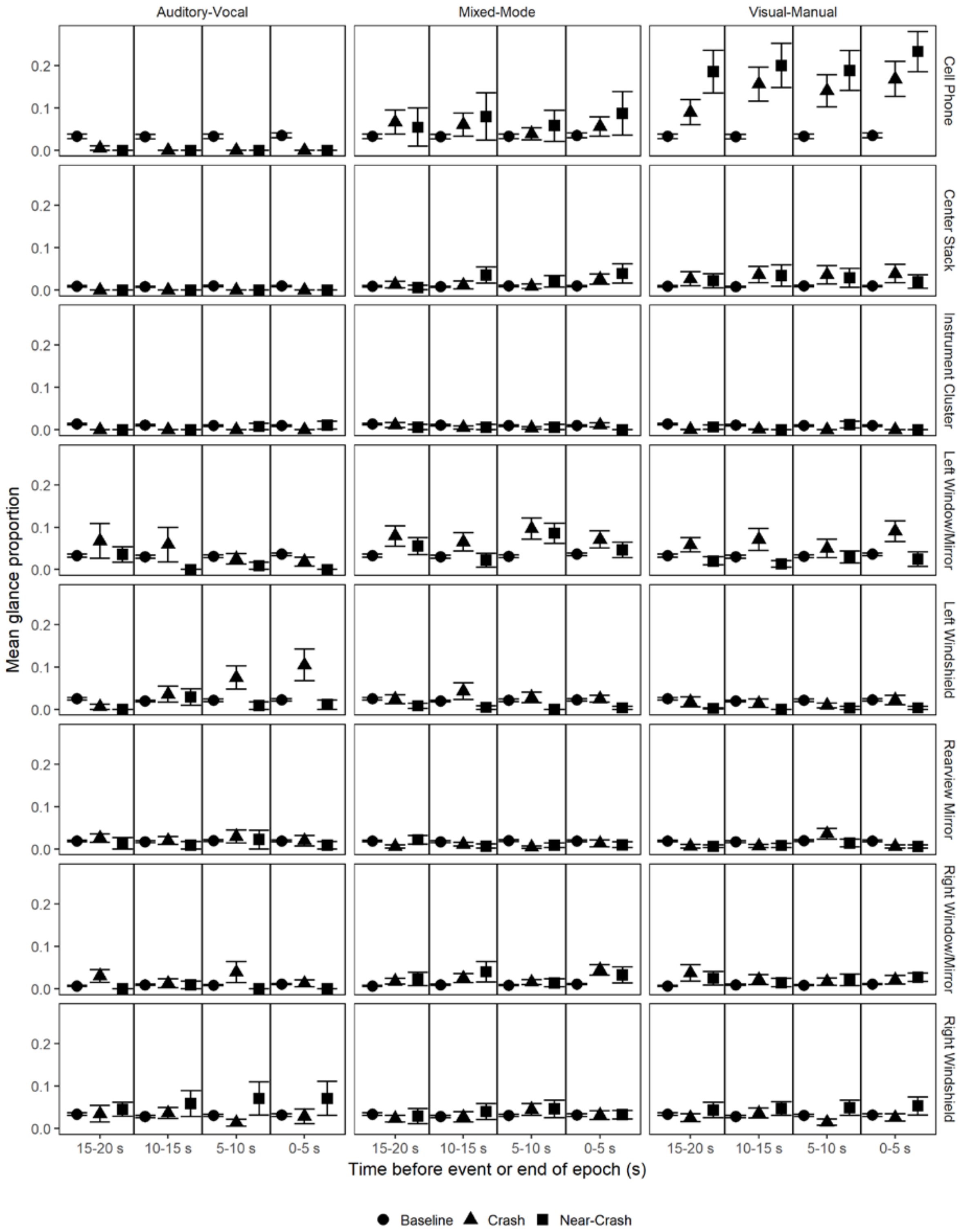

Fig. 6. *Mean glance proportion by location, task modality, SCE type and time to event. Error bars indicate standard error*

crash and Baseline and Crash and Baseline epochs containing Auditory-Vocal or Mixed-mode compositions of tasks; statistics suggest that, for SCEs containing Auditory-Vocal tasks, the trend is in the opposite direction, in the bins farthest from the precipitating events, with on-road glancing being longer in the SCE conditions than typical Baseline driving.

4. Discussion

The attention buffer provides a hybrid metric that reflects temporal patterns in how drivers allocate glances on- and off-road. The buffer concept represents information a driver can encode from the driving situation during on-road glances as well as the resulting loss of information when the driver looks away from the road. This metric produces a signal representative of how attention is managed over time and space. Statistical comparisons between SCEs and Baseline driving showed that regardless of the modality of secondary task composition, during SCEs, drivers exhibited a destabilization of attention over time not evident in Baseline driving. Further examination of these effects based on an analysis of accumulated buffer loss revealed a more pronounced fracturing of attention over time for epochs containing Visual-manual secondary task activity than those constrained to Auditory-vocal activity, evident from steeper negative slopes. These results suggest an accumulated risk in how glances are threaded over time and space when drivers deviate from how they attend to secondary tasks in Baseline driving.

Unlike patterns produced when drivers are engaged in visually-loading secondary tasks, those evident from buffer analyses of periods of performance of auditory-vocal secondary tasks indicate gaze centralization to the forward roadway. While allocation of glance to central and peripheral road regions was not accounted for in the current attention buffer implementation, the patterns produced from SCEs with auditory-vocal secondary task activity derive from long on-road glances, which have been linked to cognitive load [14, 15].

Exploration of the standard glance metrics of mean single glance duration (MSGD), number of glances, and proportion of glances to a location help to diagnose the underlying behavioural patterns produced from the buffer metric. Akin to the findings in the 100-car analysis [1], the analysis of MSGD for on- and off-road locations during SCEs indicated that, as compared to periods of baseline driving, when drivers fail to protect their ability to anticipate hazards via upstream reductions in the length of time glancing to forward roadway, they suffer a loss of awareness of the environment that disrupts how attention is managed in subsequent moments. This disruption leads to ill-timed glances off-road, reduced frequency of glances to SA-relevant locations, or to glances to inappropriate locations in the moments prior to precipitating events.

Breakdowns by task modality for these measures point to fewer, shorter glances to the forward roadway and to SA-relevant off-road locations, as well as to more frequent, longer glances to SA-irrelevant locations ascribed to the period 15-20s in advance of precipitating events for epochs that contain visually-loading secondary task activity. For those epochs that contain only auditory-vocal secondary task activity, drivers exhibited reduced sampling to both situationally-relevant left windshield and right window/mirror in the moments preceding a precipitating event, as early as 15-20s in advance of these events.

Following on from the analysis of the 100-car dataset [1], this analysis of a second naturalistic dataset provides further evidence of common patterns of attentional mismanagement in the moments prior to crashes and near-crashes that are distinctly different from periods of baseline driving. Viewed from the perspective of attention management, metrics like the attention buffer are able to produce time-history signatures of glance behaviour that reveal cumulative effects with safety-relevant implications.

5. Acknowledgment

Support for this work was provided by the Advanced Human Factors Evaluator for Automotive Demand (AHEAD) Consortium and the US DOT’s Region I New England University Transportation Center at MIT. The views and conclusions being expressed are those of the authors.

6. References

[1] Seppelt, B.D., Seaman, S., Lee, J., Angell, L., Mehler, B., Reimer, B.: ‘Glass half-full: On-road glance metrics differentiate crashes from near-crashes in the 100-car data’, Accident Analysis & Prevention, 2017, 107, pp. 48-62.

[2] Kircher, K., Ahlström, C.: ‘The impact of tunnel design and lighting on the performance of attentive and visually distracted drivers’, Accident Analysis & Prevention, 2009, 47, pp. 153–161.

[3] NHTSA, ‘Early Estimate of Motor Vehicle Traffic Fatalities for the First Nine Months (Jan-Sep) of (DOT HS 812 240)’, U.S. Department of Transportation, National Highway Traffic Safety Administration, Washington D.C., 2016.

[4] Owens, J. M., Angell, L., Hankey, J. M., Foley, J., Ebe, K.: ‘Creation of the Naturalistic Engagement in Secondary Tasks (NEST) distracted driving dataset’, Journal of Safety Research, 2015, 54, pp. 29-36.

[5] Lee, J., Sawyer, B.D., Mehler, B., Angell, L., Seppelt, B.D., Seaman, S., Reimer, B.: ‘Linking the Detection Response Task and the AttenD algorithm through the assessment of human-machine interface workload’, In Transportation Research Board 96th Annual Meeting (No. 17-06664), 2017.

[6] Seppelt, B., Seaman, S., Angell, L., Mehler, B., Reimer, B.: ‘Differentiating cognitive load using a modified version of AttenD’, In Proceedings of the 9th International ACM Conference on Automation User Interfaces and Interactive Vehicular Applications, 2017.

[7] Wolfe, B., Fridman, L., Kosovicheva, A., Seppelt, B., Mehler, B., Rosenholtz, R., & Reimer, B.: ‘Perceiving the roadway in the blink of an eye – Rapid perception of the

road environment', In Proceedings of the Driving Assessment Conference: 9th International Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, June 2017.

[8] Fridman, L., Reimer, B.L., Seaman, S., Angell, L.S., Seppelt, B.D., & Mehler, B.L.: 'What can be predicted from six seconds of driver glances?', In Proceedings of the ACM CHI Conference on Human Factors in Computing Systems, 2017.

[9] Glaser, Y. G., Llaneras, R. E., Glaser, D. S., Green, C. A.: 'Relationship between driver eyes-off-road interval and hazard detection performance under automated driving', SAE Technical Paper (No. 2016-01-1424), 2016.

[10] Samuel, S., Fisher, D. L.: 'Evaluation of the minimum forward roadway glance duration critical to latent hazard detection', In Transportation Research Board 94th Annual Meeting (No. 15-1707), 2015.

[11] Taylor, T., Masserang, K.M., Pradhan, A. K., Divekar, G., Samuel, S., Muttart, J. W., Pollatsek, A., Fisher, D. L.: 'Long term effects of hazard anticipation training on novice drivers measured on the open road', In Proceedings of the Driving Assessment Conference: 6th International Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, June 2011.

[12] Taylor, T., Pradhan, A. K., Divekar, G., Romoser, M., Muttart, J., Gomez, R., ..., Fisher, D. L.: 'The view from the road: The contribution of on-road glance-monitoring technologies to understanding driver behavior', Accident Analysis & Prevention, 2013, 58, pp. 175-186.

[13] Bates, D., Maechler, M., Bolker, B., Walker, S.: 'Fitting linear mixed-effects models using lme4', Journal of Statistical Software, 2015, 67, 1, pp. 1-48.

[14] Angell, L., Perez, M. A., Soccolich, S. A.: 'Identification of cognitive load in naturalistic driving', NSTSCE 15-UT-037, 2015.

[15] Wang, Y., Reimer, B., Dobres, J. & Mehler, B.: 'The sensitivity of different methodologies for characterizing drivers' gaze concentration under increased cognitive demand', Transportation Research Part F: Traffic Psychology and Behaviour, 2014, 26 (Part A), pp. 227-237.

7. Appendix A

Table 3 NEST tasks by SCE type

Task	Baseline			Crash			Near-crash		
	AV	MM[1]	VM	AV	MM	VM	AV	MM	VM
Adjusting/monitoring climate control	0	4	16	0	3	0	0	1	0
Adjusting/monitoring other devices integral to vehicle	0	5	10	0	4	1	0	1	3
Adjusting/monitoring radio	0	24	51	0	5	9	0	3	4
Applying make-up	0	1	1	0	0	0	0	0	2
Biting nails/cuticles	0	5	20	0	1	1	0	1	3
Brushing/flossing teeth	0	0	1	0	0	1	0	0	0
Cell phone	0	27	63	0	7	10	0	5	14
Child in adjacent seat - interaction	1	4	0	0	0	0	0	1	0
Child in rear seat - interaction	3	4	0	0	1	0	0	0	0
Combing/brushing/fixing hair	0	4	4	0	3	0	0	0	0
Dancing	0	25	8	0	1	1	0	2	1
Dialling hand-held cell phone	0	2	0	0	3	0	0	0	0
Dialling hand-held cell phone using quick keys	0	0	0	0	1	0	0	0	0
Drinking	0	7	14	0	2	1	0	0	1
Eating	0	3	15	0	2	3	0	0	2
Inserting/retrieving CD	0	1	0	0	0	1	0	1	0
Locating/reaching PDA/ other handheld device	0	1	0	0	0	0	0	0	0
Locating/reaching/answering cell phone	0	15	27	0	3	7	0	4	8
Looking at an object exter0l to the vehicle	0	32	54	0	18	11	0	11	8
Looking at pedestrian	0	1	4	0	1	1	0	0	0
Looking at previous crash or incident	0	1	0	0	0	0	0	0	0
Moving object in vehicle	0	0	2	0	2	0	0	1	0

[1] The mixed mode (MM) category was used whenever an epoch contained both visual-manual (VM) activity and an auditory-vocal (AV) activity. For example, if an epoch contained a VM activity (e.g., “looking at an object external to vehicle”) and, within the same 20s period, an AV activity took place (e.g., “conversation”), then it was classified as a MM epoch.

Object dropped by driver	0	0	0	0	0	1	0	0	1
Object in vehicle	0	16	25	0	12	7	0	1	6
Operating PDA/ other handheld device	0	1	2	0	0	0	0	0	0
Other external distraction	0	28	49	0	12	10	0	2	5
Other personal hygiene	0	9	17	0	2	4	0	1	3
Passenger in adjacent seat - interaction	107	63	0	9	23	0	5	9	0
Passenger in rear seat - interaction	12	10	0	3	4	0	0	1	0
Reaching for food- related or drink-related item	0	3	7	0	1	4	0	0	0
Reaching for object that is a manufacturer-installed device	0	1	1	0	1	0	0	0	0
Reaching for object	0	6	14	0	10	5	0	2	4
Reaching for personal body-related item	0	0	2	0	0	1	0	0	1
Reaching for, Lighting, Smoking, Extinguishing cigar, cigarette	0	8	10	0	1	2	0	1	2
Reading	0	0	0	0	1	0	0	0	0
Removing/adjusting jewellery	0	3	1	0	0	1	0	0	1
Removing/inserting/ adjusting contact lenses or glasses	0	4	3	0	1	2	0	0	0
Talking/listening on cell phone	33	13	0	11	7	0	5	2	0
Talking/singing	83	91	0	2	26	0	0	11	0
Texting on cell phone	0	15	70	0	6	15	0	4	18
Viewing PDA/ other handheld device	0	1	2	0	0	0	0	0	0
Writing	0	0	0	0	0	1	0	0	0

Design concept for a visual, vibrotactile and acoustic take-over request in a conditional automated vehicle during non-driving-related tasks

Andreas Mueller*[1], Markus Ogrizek[2], Lukas Bier[3], Bettina Abendroth[4]

[1, 2, 3, 4] Department of Mechanical Engineering, Institute of Ergonomics and Human Factors, Technische Universität Darmstadt, Darmstadt, Germany
*a.mueller@iad.tu-darmstadt.de

Abstract: Automated cars will be able to control themselves, but there will still be a need for take-over requests in critical situations that the automation system cannot handle on its own. In this paper a development and evaluation of three different take-over requests was performed. For this purpose, a total of 70 subjects took part in three independent studies conducted in a driving simulator mock-up. Within the studies three different critical scenarios with either a visual, a vibrotactile or a multimodal (combination of visual, vibrotactile and acoustic) take-over request were examined. During the automated ride, the test subjects were asked to engage in two different non-driving related tasks. The results show that all three take-over requests serve their purpose and all subjects switched from automated driving mode back to manual driving by using the steering wheel or pedals to intervene into the driving situation. Based on the results published here, a multimodal take-over request should be preferred, as it has the fastest reaction times in critical and non-critical traffic situations and consistently received good ratings within the questionnaires. A vibrotactile take-over request scored the worst in the questionnaires and participants stated that vibration as single stimulus is not being associated enough with a warning signal.

1. Introduction

Automated driving is currently one of the most discussed topics in the automotive industry. The technical development proceeds progressively and first automation systems are already available in certain driving conditions. Nevertheless, there will be situations where such systems reach their limits in conditional automated mode and will not be able to work reliably. In these cases, the driver must intervene and take over control of the vehicle as quickly as possible and with a high take-over quality.

In this manuscript, the analyzed investigation context of automated driving is based on the automation levels of SAE [20]. In addition to Manual Driving (Level 0) Assisted Driving (Level 1) exists since the adaptive cruise control system was introduced in 1998. In Partial Automation (Level 2) the vehicle autonomously assumes stabilization and the driver monitors the system at the track guidance level [6]. In Conditional Automation (Level 3) it is assumed that the driver can face away from active driving for a certain period of time and devote himself or herself entirely to non-driving related tasks (NDRT). According to the definition of SAE [20] and the NHTSA [14], the driver still has a duty in Conditional Automation to take over vehicle control within a certain period of time as requested by means of a take-over request (TOR). In this case, humans act as a fallback for the automation system.

1.1. Take-over process after automated driving

A TOR intends to generate an adequately timed response of the driver. Consequently, the driver must perceive this request explicitly. In the first step, the perception of stimuli themselves needs to be examined.

Former studies already dealt with few factors that are affecting the driver's take-over in automated driving. Radlmayr et al. [19] already proved that traffic density has a significant impact when driving in a motorway situation. Furthermore, the authors reached the conclusion that the exertion of NDRT, just as using a smartphone, worsens take-over quality in situations with heavy traffic and increases the likelihood for collisions. When showing the participants of an online survey pictures of different complex traffic situations, Eriksson et al. [7] found out that orientation occurred faster in less complex situations and when being pressed for time. Merat et al. [13] ascertained the similarity of reactions to critical incidents during automated driving without NDRT to reactions in situations during manual driving. Within the scope of a driving simulator study, Carsten et al. [3] examined the impact of three different automation levels (manual, semi-automated and highly automated) on the driver's ability to concentrate his attention on the street in association with his engagement during NDRT. Referring to this, the authors came to the conclusion that engagement in NDRT grows with a higher automation level, resulting at the same time in a decrease of the diver's focus on the street. Strand et al. [21] confirmed the negative influence of a high automation level on the driving performance after a take-over through comparing semi-automated with highly automated driving in critical situations, which occurred due to errors in the automation system. Happee et al. [9] conducted a driving simulator study that aimed on examining passing maneuvers on a motorway with blocked lanes. In the context of their research, they were able to prove a negative influence of higher automation levels on the driver's take-over as his steering and brake input occurred delayed in autonomous driving compared to manual driving. Damböck et al. [5] studied the required time needed for a take-over from autonomous driving back to manual driving in order to enable a comfortable take-over process for the driver. Based on their results the authors suggest a timeframe of at least six seconds needed for a comfortable take-over. Other studies focus on researching different sensory channels involved in a TOR during highly automated driving. Naujoks et al. [15] examined the effects of visual-auditive compared to visual-only TORs on the driver. The response time „Hands on Steering Wheel“[1] was found to be significantly shorter after a visual-auditive TOR. Petermeijer et al. [16] revealed positive functions of a vibrotactile feedback compared to an auditive TOR and the combination of both. The experiments were conducted on a simulated straightaway three-lane motorway without traffic and a driving speed of 120km/h. Within the study the drivers' response times while being involved in NDRT were evaluated. The results show that an intervention in the means of steering can be executed the fastest in a combined TOR situation. The direction of the evasive after a TOR is independent of whether the warning sound and/or the vibration was played from left or right. Petermeijer et al. [17] examined the effect of different variations of a vibrotactile TOR in a driving simulator study. The driving route was a three-lane motorway without traffic. Based on the results the participants' response times were faster when vibration was perceptible over the whole pad instead of single vibration patterns being noticeable. Telpaz et al. [22] conducted experiments with vibrotactile feedback after participants were asked to send a text message from a cell phone during autonomous driving. Within the scope of the TOR vibration was an indication for traffic. Driving took place on a simulated five-lane motorway. Response times were found to be faster for a vibrotactile TOR compared to an acoustic TOR. In summary, former studies dealt with response times dependent on the automation level, the traffic situation, NDRT and variations of a TOR.

1.2. Scope of this paper

Based on the literature, the following research question emerges: Is a unimodal TOR sufficient enough to ensure a fast reaction time between the TOR activation and the driver's intervention or does a simultaneous multimodal addressing of different sensory channels lead to better reaction times? In addition, it will be investigated whether there is a direct call to action after different TOR modalities and how disturbing the visual and acoustic TOR in particular is perceived by passengers.

For this purpose, three different TORs were developed and evaluated in three independent subject studies in this paper, see Table 1. Furthermore, reaction times between a visual, a vibrotactile and a multimodal (combination of visual, vibrotactile and acoustic) TOR will be compared.

The background for the development and evaluation is the selection of a supposedly optimal TOR. This is necessary in order to rebuild the driver's situation awareness as quickly as possible in critical driving situations. For this purpose, three realistic traffic situations, which differ from previous studies found in the literature, were developed and implemented in a driving simulator mock-up.

In Study (1) the perceived vibration intensity was examined. The scope was to identify the ideal vibration strength of a vibrotactile TOR.

In Study (2), a visual TOR was tested within three different scenarios. Test persons were asked to use their smartphones as a NDRT during the automated drive.

The vibrotactile TOR and multimodal (visual, vibrotactile and acoustic) TOR were tested for reaction times in Study (3). A tablet was offered to the subjects as a NDRT.

The aim of the NDRT is to distract the test persons as much as possible from the actual driving events and to create uniform test conditions. In addition to the objective driving data from the simulator, subjective data was collected in all studies using questionnaires. The data from Study (2) and Study (3) are compared and a design recommendation for an optimized TOR is derived from this comparison.

Table 1 Overview of the three studies used to evaluate a take-over request

	Applied studies	TOR	Test environment	NDRT	Participants	Scope of analysis
1	Vibration mat		Vehicle mockup without driving simulation	-	N = 21	Perceived Vibration
2	LED light strip	Visual	Vehicle mockup with driving simulation	Smartphone	N =19	Reaction time & subjective ratings
3	Vibration mat	Vibrotactile	Vehicle mockup with driving simulation	Tablet	N = 30	Reaction time & subjective ratings
	Vibration mat, LED light strip & acoustic warning sound	Multimodal				

[1] The Hands on Steering Wheel response time is defined as the time between the TOR entry and the first contact of hands with the steering wheel.

2. Developed take-over requests

Automated cars (Level 3 and Level 4) will be able to control themselves on different roads and in different traffic situations. However, there will still be a need for TORs in situations that the automated car cannot handle on its own as well as in planned changes of control. The TOR aims at bringing back the driver from autonomous to manual driving. Those TORs can be communicated to the driver via different modalities, as already being discussed in the introduction of this paper. Kayser et al. [12] rated the importance of different sensory channels for the vehicle guidance. For this reason, the development was reduced to a visual, acoustic and vibrotactile TOR. Within the frame of the following research, three different TORs were developed and evaluated in subject studies at the Institute of Ergonomics & Human Factors at the Technische Universität Darmstadt.

2.1. Visual take-over request via LED light strips

Research experiments have shown that LED batten luminaires have great potential as a visual warning signal compared to classic visual ADAS. Utesch [23] showed that fewer gaze averting from the road occurred, since warnings are also perceived in the peripheral field of vision. This leads to better reaction times due to selective attention theory, which says that a person can react faster to larger stimuli than to smaller ones. The LED arrangement around the driver can also be used to provide spatially oriented warnings. Common display elements used in series production, such as the combination, head-up and multimedia display cannot offer this feature.

Therefore, a visual information and warning system was developed. For this purpose, three LED light strips (LPD8806) were installed at the driving simulator mock-up, adding up to 97 individually selectable LEDs. One attached to each driver and passenger door and a third one attached to the dashboard at the height of the windscreen, see Figure 1. It was ensured that these were mounted in the driver's field of vision. The field of view of 180-200° presents a considerably greater vertical than binocular expansion (ca. 130°). Within the development and construction special attention was payed to the visibility of the LED light strips when turning away from the current street situation due to NDRTs. With the help of an Arduino microcontroller the LED light strips were dynamically controlled regarding their brightness, colour and blinking frequency and they were connected to the simulation software. In the course of the visual TOR the LEDs gave light in red and pulsated in a frequency of 2,6 Hz.

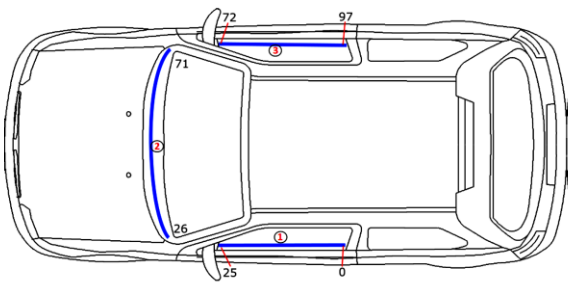

Fig. 1. *Visual LED arrangement layout with 97 LEDs*

2.2. Acoustic take-over request via loudspeakers

Acoustic signals are sensible independently of the driver's direction of view and therefore play a significant role when executing a NDRT. Furthermore, they are omnidirectional and can be perceived from every direction. Referring to Wicken's [24], theory of multiple resources, further advantages can be identified since a parallel processing of acoustic signals and visual information is possible. Regarding the selection of a suitable signal tone a study of Färber [8] was used as reference. In this study participants had to evaluate different tone frequencies in terms of urgency and amenity. Based on the results a 75 dB(A) 440Hz sinusoidal tone with a duration of one second, played every two seconds, was selected as acoustic TOR. This acoustic warning signal aims at alerting the driver without directional indication in case of a dangerous driving situation.

2.3. Vibrotactile take-over request via vibration mat

Typical visual or auditory interfaces have the disadvantage of being possibly ignored. For example, acoustic warning systems run the risk of being covered by an ambient noise or sounds of the NDRT.

The information content of vibrotactile signals is limited compared to visual or acoustic signals. However, information can be passed on to the driver independently from his field of vision and ambient sounds and will be only perceptible for himself or herself. Possible areas of application within a vehicle are the driver's seat, the back rest, the seat belt and the steering wheel. Since physical contact between the driver and the vibrating surface is essential for an information intake, certain areas of application can be classified as unsuitable. As the driver can be involved in NDRTs in autonomous driving and does not have to steer the car himself or herself, hands can be taken off the steering wheel. Petermeijer et al. [16] explained that seat belts and seats themselves are the only parts within a car that present a suitable area of application for vibrotactile feedback devices as the driver is always physically connected to them. Therefore, a vibration mat, usable within the IAD Driving Simulator as well as in non-simulators, was constructed for this research. The most useful publication to support this approach is the work of Ji et al. [11]. The authors conduct different studies, all of them referring to the intensity area of vibrotactile actuators being appropriate for human drivers and to the space needed between two actuators to feel their different localizations.

Vibration actuators that have a similar characteristic as suggested by Ji et al. [11] were used. Also, an unbalanced motor of Precision Microdrives Ltd was chosen. The relevant actuator characteristic curve of frequency and amplitude dependent on the applied voltage can be found in Figure 2. A vibration mat including 21 eccentric mass rotation actuators (Precision Microdrives 320-105) in a 7x3 arrangement was developed, see Figure 3. The mat is able to transmit both dynamic and static vibration patterns and can be used on the driver's seat in a driving simulator or in field tests. Each of the actuators can be controlled separately. Electronics and actuators were designed with focus on a wide vibration intensity spectrum. The control of each actuator is realized by means of another Arduino microcontroller with a self-developed software to ensure the connection with the simulation software.

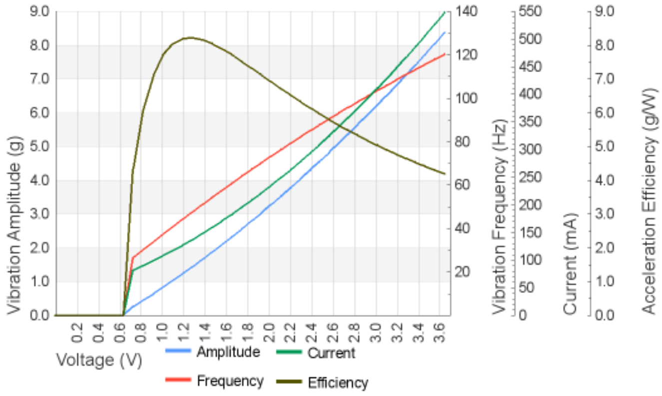

Fig. 2. *Vibration motor performance of actuator type 320-105 (Precision Microdrives [18])*

The portable vibrotactile mat consists of 2,5 cm thick foam material. Twelve cutouts in the seat back and nine cutouts in the seat cushion are made for the actuators. The actuators are situated in protecting plastic pipes which then were placed within the cutouts. The foam material features a high degree of hardness to prevent the user from sinking in and to enable a comfortable sitting. In addition to the 2,5 cm thick foam material mat with the embedded actuators, two 1 cm thick pads consisting of foam material as well and with the same degree of hardness were attached, one on top of the seat back and the other on top of the seat cushion. Those two additional pads aim at preventing the user of the vibration mat from feeling the actuators and increase the comfort of the mat.

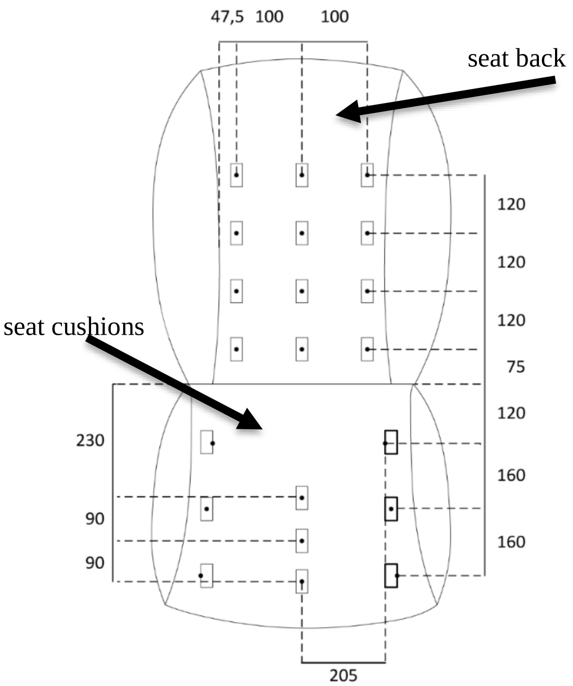

Fig. 3. *Vibrotactile mat arrangement layout with 21 (7x3) vibration motors (eccentric mass rotation). All distances are given in millimetres*

3. Method

3.1. Experimental Set-Up

Experiments were conducted in a high fidelity static driving simulator mock-up at the Institute of Ergonomics & Human Factors at the Technische Universität Darmstadt. The driving simulator consists of a full vehicle mockup (Chevrolet Aveo), a field of view of 180° front projection and a representation of all driving mirrors due to three rear projections. The simulation is realized with Silab 5.1 (WIVW) and a self-developed automation controller based on the definition of SAE [4] Level 3 Conditional Automation.

For Study (1), only the vibrotactile vibration mat was tested independently from a simulated driving task. For a realistic test environment, a total of three different critical scenarios were created which are used in Study (2) and Study (3):

3.1.1 Scenario 1 – city exit: Complete breakdown of the automation system at 50 km/h at the city exit after 110 sec of autonomous driving. As a result, the car drifts off to the right grass verge. A non-intervention of the driver leads to a collision after 3.5 sec with a street sign.

3.1.2 Scenario 2 – tunnel exit: Complete breakdown of the automation system at 100 km/h (street out of town) at the exit of a tunnel after 210 sec of autonomous driving. A non-intervention of the driver first leads to a cut into the oncoming lane and after 2.85 sec to a collision with a reflector post and a couple of trees.

3.1.2 Scenario 3 – broken-down vehicle: TOR during an inner-city left turn at 50 km/h after 350 sec of autonomous driving due to a broken down vehicle on the same lane. Breakdown of the longitudinal control, lane and speed stay constant. A non-intervention of the driver leads to an accident after 5.8 seconds.

Overall, three different TORs will be analyzed within this paper: a visual TOR, a vibrotactile TOR and a multimodal (combination of visual, vibrotactile and acoustic) TOR. The three different scenarios and TORs form a 3x3 experimental design. Between the vibrotactile TOR and the multimodal TOR as well as between a visual TOR and a vibrotactile TOR a within-subject design was chosen. Between a visual TOR and a multimodal TOR, a between-group design was set. In order to minimize the positive effect of learning on the driver’s take-over reaction, the subject group tested the respective scenarios and TORs in permuted order in conditional automation mode according to SAE [20] Level 3. In all scenarios, the driver has to switch from the NDRT to traditional manual driving. 20 sec after the successful take-over the simulation paused and participants had to answer a questionnaire. Afterwards the next scenario followed. Altogether the driving simulator experiment had a duration of 30 min in Study (2) and 60 min in Study (3).

3.2. Examined parameters

A questionnaire regarding an evaluation on a 7-Point Likert Scale (very pleasant – very unpleasant) of the perceived vibration was given to the participants of Study (1). Questionnaires were originally written in German and have been translated afterwards into English for the purpose of this paper. The participants had to evaluate twelve different vibration strengths, given in permuted order for 5 sec each, one after another.

In order to compare the different TORs in Study (2) and Study (3) with each other, subjective and objective measures were collected during each of the studies. Objective driving data from the simulator was recorded.

The subject's reaction time to a TOR is characterized as the period of time from the moment the simulation software started the TOR to the moment the driver reacts to it and intervenes into driving. Intervention could happen in the form of actuating at least one of the classic vehicle controls steering wheel and pedal. Intervention through the steering wheel is captured starting at a change in angle of 2°. Intervention through operating the gas or brake pedal is detected when the pedal position changes by more than 10% from its initial position. The intervention, which was first made by the driver, will be considered as a minimum reaction time in the further process. A self-developed questionnaire was distributed to the subjects after every TOR. With the questionnaire subjects assessed the perceived urgency, usability, distraction and comfort transmitted by the TOR. All questions were asked in German and the participants were able to rate the TOR on a 7-Point Likert Scale.

3.3. Execution of non-driving related tasks before take-over request

During the automated ride, the subjects were asked to engage in NDRTs. In advance of the actual test execution participants were given an explanation of the functional principles of an automatized driving car. Thereby, the subjects were also given the information that a focus of one's attention on the driving situation was not necessary anymore and that an occupation in a NDRT was possible instead.

In order to attain an equal degree of distraction and a consistent experimental design across all participants, possible NDRTs were selected in advance. When investigating the visual TOR (Study 2), the subjects were asked to actively distract themselves from the driving activity and to interact with their own smartphone.

As the participants of Study (2) did not use their smartphone during the entire automated ride, participants of Study (3) (vibrotactile and multimodal TOR) were asked to complete a cognitively demanding test on a tablet (Huawai MateBookE). For this purpose the Brain Workshop program was installed on the tablet [2]. This program is based on a n-back test, used as a dual 2-back test within the study. Hereof, a blue visual stimulus is presented in random order in a 3x3 matrix. At the same time a letter is announced acoustically with every new presentation of the blue stimulus. With every new presentation and announcement, the subject has to identify if the forelast (2-back) stimulus and letter combination is congruent to the current one. If a repetition is detected correctly, a button on the tablet must be tapped accordingly, depending on the stimulus. The test's goal is to identify as many congruent pairs as possible and the study's participants have therefore been motivated to perform best possible. A more precise explanation of the n-back task can be found in [10].

3.4. Subject studies

The results of this paper are based on three independent subject studies. Participants have been acquired via notices at the TU Darmstadt and a subject database. In Study (1) (evaluation of the vibration intensity) 21 people, six of them women, participated. The subjects' average age was 27.3 years (SD 9.5 years). In Study (2) (visual TOR) 19 people, five of them women, participated (MN = 24.7 years, SD = 5.7 years). 30 people, eight of them women, participated in Study (3) (vibrotactile and multimodal TOR) The subjects' average age was 33.2 years (SD 6.8 years). In all three studies participants did not have any former experience with highly automated driving simulator mock-ups.

3.5. Statistical Evaluation

The parameters examined are displayed in a BoxPlot diagram, indicating the arithmetic mean [1]. Different test procedures are used for the statistical evaluation. The significance level is set to $\alpha = 0.05$. For the testing for standard distribution the Shapiro-Wilk test is used. As far as a standard distribution of the two samples is present, a T-test for dependent samples is performed in Study (3). For comparing Study (2) and Study (3) with each other a T-test for independent samples is used. Thereby, homogeneity of variance is examined with the Levene-test. As a result of the mean value comparison, a prediction can be made as to whether the considered mean values differ significantly from each other ($p \leq 0.05$) or not. If the results differ significantly, the effect strength is calculated according to Cohen [4].

4. Results

The results of the studies are presented in the following section. The descriptive data as well as the interference statistics will be mentioned as well.

4.1. Perceived Vibration

The aim of the first study was to adjust the constructed and built vibration mat to an optimized vibration intensity. The structure of the study and the number of required participants is based on Ji et al. [11]. Twelve different vibration intensities were transmitted in permuted order to the subjects in the simulator mock-up. The subject then had to evaluate each level on a Likert scale from one (very unpleasant) to seven (very pleasant). Any level of vibration intensity is held for five sec. Between each level there is a pause of five sec for the participant to complete the questionnaire.

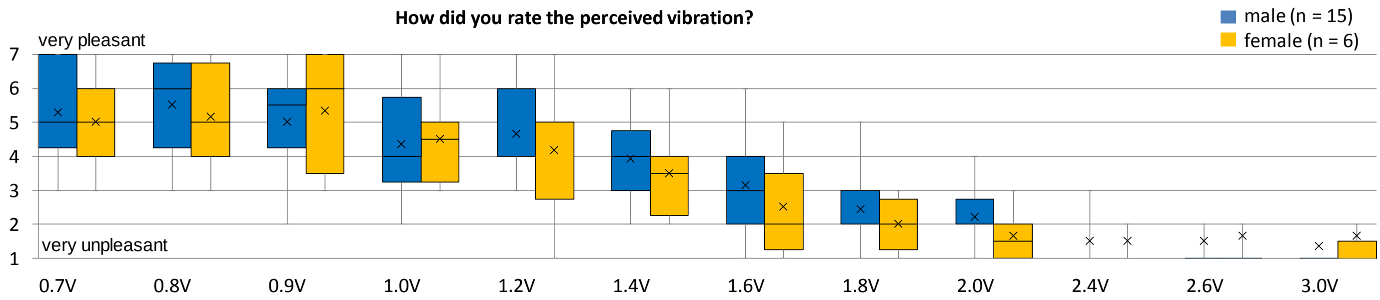

Fig. 4. *Rated vibration intensity as a function of gender and applied voltage*

Table 2 *Dataset: rated vibration intensity as a function of gender and applied voltage*

Voltage [V]	MN ♂ / ♀	SD ♂ / ♀	Voltage [V]	MN ♂ / ♀	SD ♂ / ♀
0.7	5.29 / 5.00	1.49 / 1.55	1.6	3.14 / 2.50	1.41 / 1.64
0.8	5.50 / 5.17	1.34 / 1.72	1.8	2.43 / 2.00	1.22 / 0.89
0.9	5.00 / 5.33	1.66 / 1.97	2.0	2.21 / 1.67	0.80 / 0.82
1.0	4.36 / 4.50	1.74 / 1.52	2.4	1.50 / 1.50	0.94 / 0.84
1.2	4.64 / 4.17	1.34 / 2.23	2.8	1.50 / 1.67	0.65 / 0.82
1.4	3.93 / 2.50	1.41 / 1.64	3.0	1.36 / 1.67	0.50 / 1.03

As seen in Figure 4 and Table 2, the feeling of pleasure decreases with increasing vibration intensity. Low vibration levels (0.7 V / 30 Hz / 0.3g) are rated as pleasant and high vibration levels (3V / 105 Hz / 6g) as very unpleasant. As also described in Ji et al. [11] a gender dependence on the perceived sensation of vibration was observed. Female test subjects rated the vibration intensity in the range of 1.2 V to 2.0 V as more unpleasant compared to male test subjects. With low and high vibration intensities, there are hardly any gender differences. Due to the low number of test persons, interference statistics were dispensed. According to the results of Study (1), a value of 1.4 V was chosen for the selection of the optimal vibration intensity, as this was evaluated by the test persons as the average between very pleasant and very unpleasant. Based on this, the conclusion can be drawn that in further studies subjects neither will not notice the vibration due to a too low intensity nor will they be distracted too much by an excessively vibration intensity set.

4.2. Reaction times between a take-over request and the driver's intervention

In the following, the reaction times from Study (2) and Study (3) between the initiated TOR and the driver's intervention are summarized and explained.

In scenario 1, after the vehicle has been driven through a city in a conditional automated mode for 110 sec, the automation controller fails at the city exit and the TOR is activated. In this scenario, there were no significant differences between the three different developed TORs. However, it turns out that in the case of the visual TOR participants require the longest period of time ($MN_{visual,S1}$ = 1.55 sec, $SD_{visual,S1}$ = 0.87 sec, n = 14) to intervene after the TOR activation. The fastest response times were observed with the multimodal TOR ($MN_{multimodal,S1}$ = 1.12 sec; $SD_{multimodal,S1}$ = 0.21 sec, n = 10), followed by the vibrotactile TOR ($MN_{vibrotactile,S1}$ = 1.36 sec, $SD_{vibrotactile,S1}$ = 0.43 sec, n = 25). Participants who were requested to resume to manual driving by a vibrotactile TOR ($MN_{vibrotactile,S2}$ = 1.35 sec; $SD_{vibrotactile,S2}$ = 0.39 sec, n = 22) in scenario 2 (omission of road markings), took over significantly faster compared to when the TOR was transmitted visually ($MN_{visual,S2}$ = 2.05 sec; $SD_{visual,S2}$ = 0.81 sec, n = 18, t(23,406) = 3.389, p = .002). The effect strength according to Cohen [4] is d = .57 and corresponds to a medium effect. An even greater effect strength can be seen when comparing the visual TOR with the multimodal TOR ($MN_{multimodal,S2}$ = 1.14 sec; $SD_{multimodal,S2}$ = 0.22 sec, n = 23, t(18.932) = 4.644, p = .000, d = .73.

In scenario 3, where the automation controller does not clearly recognize a broken down vehicle in the city and starts the TOR approximately six sec before the imminent collision, similar results compared to those in scenario 2 can be found. In the case of a visual TOR ($MN_{visual,S3}$ = 1.95 sec, $SD_{visual,S3}$ = 0.62 sec, n = 18) the subjects intervene significantly later than in the case of a vibrotactile TOR ($MN_{vibrotactile,S3}$ = 1.46 sec, $SD_{vibrotactile,S3}$ = 0.36 sec, n = 21, t(37) = 3.024, p = .005, d = .45). Response times for a multimodal TOR ($MN_{multimodal,S3}$ = 1.22 sec, $SD_{multimodal,S3}$ = 0.23 sec, n = 20) prove to be significantly faster than for the visual TOR, t(21.070) = 4.729, p = .000, d = .72, and significantly faster in comparison to the vibrotactile TOR, t(14) = 2.446, p = .028, d = .55.

At the end, the reaction times between the start of the TOR and the first measurable driver intervention were averaged over all three scenarios. Similar to the results of scenario 3, significant differences between the visual TOR ($MN_{visual,av.}$ = 1.89 sec, $SD_{visual,av.}$ = 0.60 sec, n = 19) and the vibrotactile TOR ($MN_{vibrotactile,av.}$ = 1.39 sec, $SD_{vibrotactile,av.}$ = 0.27 sec, n = 29, t(22.962) = 3.451, p = .002, d = .58) can be observed. Response times for the multimodal TOR ($MN_{multimodal,av.}$ = 1.17 sec, $SD_{multimodal,av.}$ = 0.20 sec, n = 27) prove to be significantly faster on average than for the visual TOR, t(20.835) = 5.055, p = .000, d = .74, and significantly faster

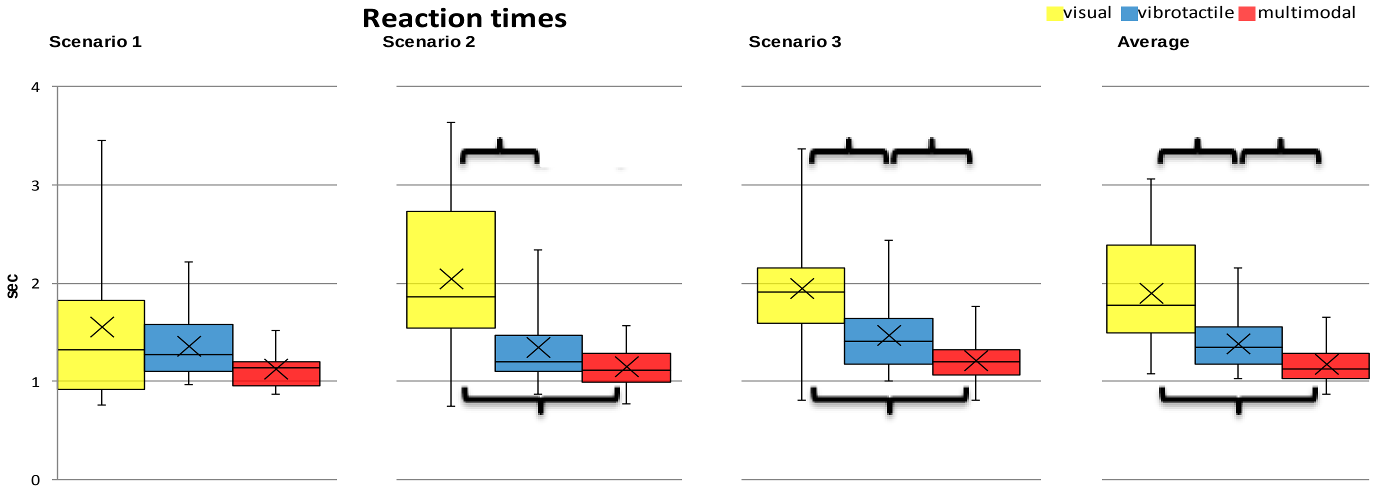

Fig. 5. *Reaction times between the examined TOR and the driver intervention depending on the scenario*

Table 3 *Dataset: Reaction times between the examined TOR and the driver intervention depending on the scenario*

TOR	Scenario 1 MN / SD [sec]	Scenario 2 MN / SD [sec]	Scenario 3 MN / SD [sec]	Average MN / SD [sec]
Visual	1.55 / 0.87	2.05 / 0.81	1.95 / 0 62	1.89 / 0.60
Vibrotatcile	1.36 / 0.34	1.35 / 0.39	1.46 / 0.37	1.39 / 0.27
Multimodal	1.12 / 0.21	1.14 / 0.22	1.22 / 0.23	1.17 / 0.20

than for the vibrotactile TOR, t (26) = 5.215, p = .000, d = .72. The results of the reaction times are shown in Table 3 and Figure 5 using box plots.

4.3. Questionnaires

In addition to the objective driving data, a questionnaire was handed out to the subjects after each TOR. The results of the subjective survey are shown in Figure 6 and Table 4. According to this, most subjects from Study (2) perceived the visual TOR to be urgent, but not very urgent ($MN_{visual,Q1}$ = 4.89 , $SD_{visual,Q1}$ = 1.28, n = 19). There is a difference in the ratings of the vibrotactile feedback ($MN_{vibrotactile,Q1}$ = 4.03 , $SD_{vibrotactile,Q1}$ = 1.17, n = 30). This was evaluated significantly less urgently than the visual TOR, t(47) = 2.44, p = .019, d = .34. The multimodal TOR was most urgently assessed by the subjects ($MN_{multimodal,Q1}$ = 5.02, $SD_{multimodal,Q1}$ = 0.96, n = 30) and differs significantly from the vibrotactile TOR, t(29) = -4.05, p = .000, d = .60.

Regarding the second question, subjects assessed the different TORs according to their usefulness. The visual TOR ($MN_{visual,Q2}$ = 5.21 , $SD_{visual,Q2}$ = 1.28, n = 19) tended to be more useful than the vibrotactile TOR ($MN_{vibrotactile,Q2}$ = 4.69 , $SD_{vibrotactile,Q2}$ = 1.36, n = 30). However, a significant difference in usefulness evaluation could only be determined between the vibrotactile TOR and the multimodal TOR ($MN_{multimodal,Q2}$ = 5.25, $SD_{multimodal,Q2}$ = 1.33, n = 30), t(29) = -2.446, p = .021, d = .41.

It can be seen that the majority of the test persons did not perceive the warning system as disturbing. Between the visual ($MN_{visual,Q3}$ = 5.74 , $SD_{visual,Q3}$ = 1.29, n = 19), the vibrotactile ($MN_{vibrotactile,Q3}$ = 5.63 , $SD_{vibrotactile,Q3}$ = 1.18, n = 30) and the multimodal TOR ($MN_{multimodal,Q3}$ = 5.41, $SD_{multimodal,Q3}$ = 1.40, n = 30) were no significant differences.

The perceived comfort is tending to be the highest with the visual TOR ($MN_{visual,Q4}$ = 5.08 , $SD_{visual,Q4}$ = 1.25, n = 18) but no significant differences can be found between the different variants ($MN_{vibrotactile,Q4}$ = 4.79 , $SD_{vibrotactile,Q4}$ = 1.13, n = 30; $MN_{multimodal,Q4}$ = 4.53, $SD_{multimodal,Q4}$ = 1.22, n = 30.).

Finally, all three TORs were generally judged on a 7-Point Likert Scale (recommend - not recommend). The visual TOR ($MN_{visual,Q5}$ = 5.63 , $SD_{visual,Q5}$ = 1.09, n = 16) was recommended significantly more often than the vibrotactile TOR ($MN_{vibrotactile,Q5}$ = 4.57 , $SD_{vibrotactile,Q5}$ = 1.91, n = 30, t(43.716) = 2.396, p = .021, d = .34). On average, the visual and the multimodal TOR ($MN_{multimodal,Q5}$ = 5.60, $SD_{multimodal,Q5}$ = 1.90, n = 30) hardly differ from each other and no significant difference to the vibrotactile TOR was found.

5. Discussion

The discussion is divided into three sections. First, the design of the individual TORs is critically questioned and frequently mentioned statements by the study's participants are mentioned. In the further course of the discussion, the determined reaction times are compared with each other and with different literature references. Furthermore, the questionnaire data will be discussed. Finally, the research questions from chapter 1.2 will be discussed and answered as well.

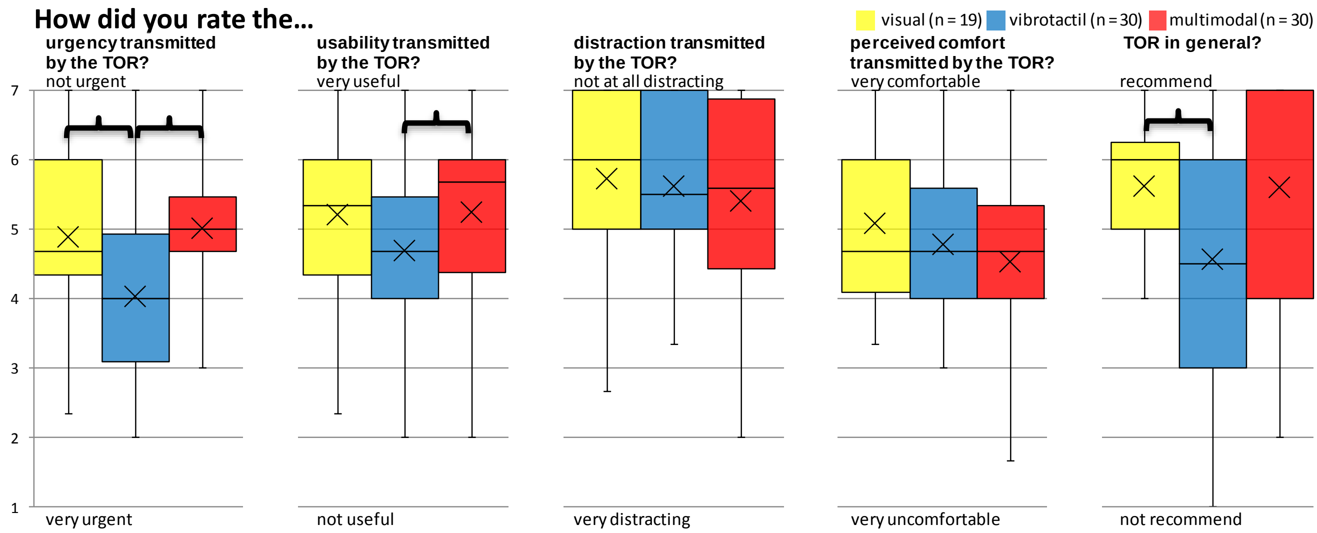

Fig. 6. *Subjective ratings of examined TOR*

Table 4 *Dataset: Subjective ratings of examined TOR*

TOR	Urgency MN/SD	Usability MN/SD	Distraction MN/SD	Comfort MN/SD	In General MN/SD
Visual	4.89 / 1.28	5.21 / 1.28	5.74 / 1.29	5.08 / 1.25	5.63 / 1.09
Vibrotatcile	4.03 / 1.17	4.69 / 1.36	5.63 / 1.18	4.79 / 1.13	4.57 / 1.91
Multimodal	5.02 / 0.96	5.25 / 1.33	5.41 / 1.40	4.53 / 1.22	5.60 / 1.90

5.1. TOR design concept

It turns out that the visual TOR, due to its alarming red pulse frequency, is intuitively understandable and is well suited as a warning system. Despite carrying out a NDRT and therefore turning eyes away from the road the system with the visual TOR is still well visible in the driver's peripheral field of vision. Test subjects from Study (2) testified that they would support an audible warning in addition to the visual stimulus. In principle, the test persons did not find the system disturbing. Only reflection effects on the mock-up's windscreen were noted negatively.

The acoustic TOR was positively perceived by the subjects in Study (3). The warning tone of 440 Hz was noticed by all participants, despite simulated traffic and wind noise. Overhearing of the warning signal, even when a NDRT is executed, did not occur due to the volume of 75 db (A). In the performed study, the acoustic signal was only tested in combination with the vibrotactile and the visual TOR. Whether an acoustic TOR leads to different reaction times should be investigated in another study.

Since hardly any results about the required vibration intensity were available in the literature, a vibration intensity recommendation was determined based on the results of Study (1). For the used actuators (Precision Dynamics 320-105), this is a vibration frequency of approx. 50 Hz and a vibration amplitude of 1.5 g at an applied voltage of 1.4 V. This value was chosen because it presents the average between "very pleasant" and "very unpleasant" rated by the test persons.

It can be assumed that in further tests, subjects neither will not perceive the vibration due to a too low vibration intensity nor will they be frightened by a too high vibration intensity, which would lead to a poorer take-over quality.

It is also noteworthy that there were differences at the perceived vibration intensity in terms of gender. Female subjects evaluated the perceived vibration intensity more unpleasant than male subjects, especially in the medium voltage range (1,2 – 2,0 V). Similar results were shown in Ji et al. [11] and can be confirmed by this study. Whether this effect actually depends on gender or body weight should however be examined in further studies. Individual test persons point out that various areas of vibration were perceived as very unpleasant. The entries vary from subject to subject, so that no generally valid statement can be made. Nevertheless, the back area in general and the kidney area in particular are more frequently mentioned.

Furthermore, based on the test person's evaluations the visual red light bar and the loud warning tones appear threatening and can also alarm and frighten the passengers. The vibrotactile feedback, on the other hand, is very private and can only be perceived by the contact person.

5.2. Reaction times

One of the most important criteria for the evaluation of a TOR are the reaction times, which need to be as short as possible in critical real driving situations.

The experiments within this study show that under the same scenarios, the fastest reaction times are caused by the multimodal TOR, followed by the vibrotactile and the visual TOR. Especially in scenarios 2 and 3 these differences are significant and show high effect strengths.

A possible reason why the visual TOR led to delayed reaction times could be due to the fact that the subjects were occupied by a visual NDRT. A visual stimulus right before the activation of a visual TOR could therefore result in a delay of information processing and ultimately in the execution of an action. This also speaks for the multiple resource theory according to Wickens [24]. Whether similar effects occur in the case of an acoustic NDRT being performed right before the activation of an acoustic TOR should be further investigated.

Furthermore, the different response times may have been caused by the different NDRTs used in Study (2) and Study (3). In Study (2) subjects held their private smartphone in their hands with the incitement to do everyday things. Since not all subjects operated the smartphone continuously during the experiment, a tablet was attached to the central information display position in Study (3) and a cognitively highly demanding dual 2-back test needed to be executed by the subjects. Despite the supposedly higher cognitive demand, the reaction times are significantly shorter. One possible cause could be that in the case of a TOR the test persons did not want to drop their smartphone directly out of their hands and tried to put it down safely.

The influencing factor of the time budget and indirectly of the take-over situation's criticality as well, which has already been investigated by Damböck [5], could also be found in this study. In this case, scenario 3 was the most uncritical, as it specified a time budget of approximately six secs before a collision with a broken down vehicle occurs. The results show that in scenario 3 the reaction times were longer than in scenario 2, where the time budget amounts for only approximately 3.5 sec. Furthermore, scenarios 2 and 3 differ from the failure of the automation controller. While in scenario 2 (and 1 as well) the automation system fails completely, in scenario 3 only the longitudinal controller was deactivated. Scenario 3 is therefore less critical, as the vehicle does not drift off the road. Ultimately, it can be concluded that the less critical the situation, the longer the reaction times.

A comparison of reaction times with literature data shows that the take-over times after a vibrotactile TOR found by Petermeijer et al. [16] and Petermeijer et al. [17] are approximately 2.67 sec and 1.97 sec respectively. These values are considerably slower than the results of this study ($MN_{visual,av.}$ = 1.89 sec, $MN_{vibrotactile,av.}$ = 1.39 sec, $MN_{multimodal,av.}$ = 1.17 sec).

5.3. Questionnaires

The subjective data from the questionnaires reinforce the previously described results from chapter 5.2. The subjects expressed that the visual TOR effectively informs about the need for intervention since a direct call to action regarding the relevant area is established, in this case the windscreen and thus the external traffic events. In the case of the vibrotactile TOR, the vibration stimulus is not associated with an operational intent and test persons often did not know exactly what to do. A warning effect by a vibrotactile TOR is therefore not guaranteed; this can also be confirmed by the question of urgency.

The results regarding the question about the perceived usability confirms the aforementioned thesis that a vibrotactile signal may perceived as unhelpful. The perceived distraction does not differ from the three variants and is not perceived as disturbing by the majority of the subjects.

Although Study (1) determined a trade-off for the best vibration intensity, the questionnaire results show that the perceived comfort for a vibrotactile stimulation was rated lower than for a visual TOR.

6. Conclusion

The results show that all three TORs serve their purpose and all participants switched from automated driving mode back to manual driving by using the steering wheel or pedals again. For further investigations, in cases where the time aspect of the TOR is decisive, a TOR should be used, that causes the shortest reaction times. Three different TOR variants were developed based on existing literature results. Furthermore, the TORs were tested for their reaction times in three different critical scenarios and were subjectively evaluated using questionnaires. A total of 70 subjects took part in the three independent studies.

The vibrotactile TOR scored the worst in the questionnaires, since the vibration stimulus appears to be not associated enough with a warning signal. The different NDRT or parallel processing of the visual channel can explain the higher response times of the visual warning system.

Based on the results published here, a multimodal TOR should be preferred as it implies the fastest reaction times in critical and non-critical traffic situations as well as it has consistently good ratings from the questionnaires. Future studies should continue with the parameterization of individual factors of the multimodal TOR and optimize them with a uniform test design.

7. Acknowledgments

The authors gratefully acknowledge the following people who have contributed to this paper with their engagement: Deußen, Benjamin; Eggert, Lukas; Hadamik, Katrin Juliane; Keshavarzi, Arash; Mende, Hartmut; Moske, Jonas; Pfliegensdörfer, Felix; Romana, Simranjeet Singh; Simon, Marcel; Weiler, Henrik Paul.

8. References

[1] Bortz, J. and Weber, R. Heidelberg : Springer Berlin Heidelberg, 2005. *Statistik. Für Human- und Sozialwissenschaftler von Jürgen Bortz, René Weber*. Springer-Lehrbuch. Springer Berlin Heidelberg, Berlin, Heidelberg.

[2] *Brain Workshop. Dual N-Back game. http://brainworkshop.sourceforge.net/*.

[3] Carsten, O., Lai, F. C. H., Barnard, Y., Jamson, A. H., and Merat, N. 2012. Control task substitution in semiautomated driving. Does it matter what aspects are automated? *Human Factors: The Journal of the Human Factors and Ergonomics Society* 54, 5, 747–761.

[4] Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Taylor and Francis, Hoboken.

[5] Damböck, D., Farid, M., Tönert, L., and Bengler, K. 2012. Übernahmezeiten beim hochautomatisierten Fahren. *Tagung Fahrerassistenz. München*.

[6] Donges, E. 1982. Aspekte der aktiven Sicherheit bei der Führung von Personenkraftwagen. *AUTOMOB-IND* 27, 2.
[7] Eriksson, A., Marcos, I. S., Kircher, K., Västfjäll, D., and Stanton, N. A. 2015. The Development of a Method to Assess the Effects of Traffic Situation and Time Pressure on Driver Information Preferences. In *Engineering Psychology and Cognitive Ergonomics*, D. Harris, Ed. Lecture Notes in Computer Science. Springer International Publishing, Cham, 3–12. DOI=10.1007/978-3-319-20373-7_1.
[8] Färber, B. 1987. *Geteilte Aufmerksamkeit. Grundlagen und Anwendung im motorisierten Strassenverkehr*. Mensch, Fahrzeug, Umwelt 20. Verl. TÜV Rheinland, Köln.
[9] Happee, R., Gold, C., Radlmayr, J., Hergeth, S., and Bengler, K. 2017. Take-over performance in evasive manoeuvres. *Accident; analysis and prevention* 106, 211–222.
[10] Jaeggi, S. M., Buschkuehl, M., Jonides, J., and Perrig, W. J. 2008. Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America* 105, 19, 6829–6833.
[11] Ji, Y. G., Lee, K., and Hwang, W. 2011. Haptic perceptions in the vehicle seat. *Hum. Factors Man.* 21, 3, 305–325.
[12] Kayser, H. J., Sanders, A. F., Hess, M., Spijkers, W. A. C., and Bartmann. 1989. *Das Wahrnehmungsverhalten des Kraftfahrers in Abhängigkeit von der gefahrenen Geschwindigkeit und der Straßenraumgestaltung :Forschungsbericht aus dem Forschungsprogramm des Bundesministers für Verkehr und der Forschungsgesellschaft für Straßen- und Verkehrswesen E.V.* Bundesminister für Verkehr, Bonn-Bad Godesberg.
[13] Merat, N., Jamson, H., Lai, F., and Carsten, O. 2012. Highly Automated Driving, Secondary Task Performance, and Driver State. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 54, 5, 762–771.
[14] National Highway Traffic Safety Administration. 2012. *Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices* NHTSA-2010-0053. NHTSA.
[15] Naujoks, F., Mai, C., and Neukum, A. 2014. The effect of urgency of take-over requestsduring highly automated driving underdistraction conditions. In *Advances in human factors and ergonomics 2014. 5th International Conference on Applied Human Factors and Ergonomics : 20 volume set: proceedings of the 5th AHFE Conference, 19 - 23 July 2014*, R.-L. Jang and T. Ahram, Eds. AHFE Conference.
[16] Petermeijer, S., Bazilinskyy, P., Bengler, K., and Winter, J. de. 2017. Take-over again. Investigating multimodal and directional TORs to get the driver back into the loop. *Applied ergonomics* 62, 204–215.
[17] Petermeijer, S. M., Cieler, S., and Winter, J. C. F. de. 2017. Comparing spatially static and dynamic vibrotactile take-over requests in the driver seat. *Accident; analysis and prevention* 99, Pt A, 218–227.
[18] Precision Microdrives TM. 2018. *vibriation motor performance type 320-105*. https://www.precisionmicrodrives.com/product/320-105-20mm-vibration-motor-25mm-type. Accessed 14 May 2018.
[19] Radlmayr, J., Gold, C., Lorenz, L., Farid, M., and Bengler, K. 2014. How Traffic Situations and Non-Driving Related Tasks Affect the Take-Over Quality in Highly Automated Driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 58, 1, 2063–2067.
[20] SAE J3016. 2014. *Automated Driving. Levels of driving automation are defined in new SAE International Standard J3016*.
[21] Strand, N., Nilsson, J., Karlsson, I. M., and Nilsson, L. 2014. Semi-automated versus highly automated driving in critical situations caused by automation failures. *Transportation Research Part F: Traffic Psychology and Behaviour* 27, 218–228.
[22] Telpaz, A., Rhindress, B., Zelman, I., and Tsimhoni, O. 2017. Using a Vibrotactile Seat for Facilitating the Handover of Control during Automated Driving. *International Journal of Mobile Human Computer Interaction* 9, 3, 17–33.
[23] Utesch, F. *Unscharfe Warnungen im Kraftfahrzeug. Eignet sich eine LED-Leiste als Anzeige für Fahrerassistenzsysteme?* Dissertation. Technische Universität Carolo-Wilhelmina. Braunschweig.
[24] Wickens, C. D. 2008. Multiple resources and mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 50, 3, 449–455.

Change the way to manage an in-vehicle menu selection and thereby lower cognitive workload?

Nikolai Pärsch [1*], Martin Baumann [2], Arnd Engeln [3], Lutz W.H. Krauß [1]

[1] UI/UX Display & Interaction, Dr. Ing. h.c. F. Porsche AG, Germany
[2] Institute of Psychology and Education Dept. Human Factors, Ulm University, Germany
[3] Interaction Design and User Experience (IDUX), Stuttgart Media University, Germany
[*] nikolai.paersch@porsche.de

Abstract: The research reported here aims to investigate in more detail cognitive workload of in-vehicle information systems (IVIS). Various operating concepts for one specific task are tested. In detail, a function selection is implemented as a hierarchical menu selection or as a search function with text input. The text input modes are varied between speech, touch keyboard and touch gesture (handwriting) and tested with a driving simulator study. Main findings are, that cognitive workload of search via speech input is lower than the other alternatives being tested. Compared to workload of n-back levels, speech input is lower and manual interactions have a cognitive workload that is comparable to a level between 1- and 2-back tasks. Training effects are mainly observed at menu selection as well as text input by handwriting. Impact of operating errors on cognitive workload seems to be high and should be researched in further studies.

1. Introduction

There are several alternatives to manage a function selection within the car for designers of IVIS. Those alternatives may be divided into two groups: a selection via hierarchical menu structure or a selection via key word search and text input. For users each version seems to has its pros and cons. Lee [1] explains the main advantages of a hierarchical menu: the states of the program are displayed explicitly, so the action for the user is more recognition than recall. Search based functions on the other hand, don't need users to adapt to a certain logic but also require them to have specific keywords in mind.

The suitability of these systems for the driving context has been examined very often by measuring their *visual* workload (see e.g. Heinrich [2]). Today, there are approaches that also consider *cognitive* workload in the vehicle. Strayer et al. [3] show, that cognitive workload varies depending on task type (e.g. calling) or the mode of interaction (center stack, auditory vocal, center console).

Concerning menu and search design alternatives, there is lot of research, which should be summarized in the following lines.

1.1. Menu-Driven Systems

Hierarchical menus have a long history in computer systems. Lee [1] defines menus as user-selectable data. These can be found in most technical products with graphical user interface, also in passenger vehicles. Many guidelines exist about relevant factors for designing a good hierarchical menu system. Some of the factors are stated in Norman [4], for example: "depth versus breadth", "organization of lists", "clustering" and "item meaningfulness and distinctiveness". These design factors for menus could have implications to the visual workload needed (e.g. Burnett et al. [5], Hornof et al. [6]) but also on the cognitive workload (Matsuo [7]).

Depth versus breadth addresses the question, how many items should be displayed at one page and how many pages should follow on the next layers. Burnett et al. [5] evaluate different combinations and show, that at structured menus (arranged alphabetically), breadth is favored over depth. For unstructured menus (arranged randomly), that finding applies conversely.

Organization of lists addresses the order of list entries. The adequate ordering method may differ depending on the specific use case. In short, there is alphabetic, numeric, chronological, cognitive, semantic and an ordering by frequency of use.

Clustering means the organization of list items. To find an optimal list there are two ways to cluster content: top-down or bottom-up. Regarding top-down, the designer starts with first-order categories and divides the entries step by step until he arrives at the last level. Bottom-up the designer looks at all items and clusters them by similarity, then groups them step by step into larger groups until all the groups are combined.

Item meaningfulness and distinctiveness concerns the verbalization of items. Items should transfer information but also should be distinct to each other. As an addition the use of graphics e.g. icons could be suitable to solve these issues at some points.

1.2. Search-Driven systems

Search driven concepts are also well-known in technical systems, e.g. the probably most common example: the google search. Users provide keywords to retrieve their desired information or get to the desired stage in the interactive system. The search function always consists of a text input, where keywords could be typed in. The way of carrying out the text input, often differs between the context of use and operating device. In the vehicle the most common types of text input are done by rotating wheel, by touchscreen keyboard, by handwriting gesture on a touchpad as well as a text input via speech.

Graf et al. [8] show the suitability for this kind of search function in the context of IVIS. They compare two kinds of search functions: a quick search, where users can freely type in search terms and a categorical search, where users narrow down their search results by choosing a corresponding category. They analyze, that the search approach seems to be equally suitable or even superior to menu driven interaction. But how about different ways to carry out a text input? This comparison is made for instance in Kujala et al. [9]. They compare touch keyboard, handwriting and text input by voice recognition by their workload. Voice recognition shows the lowest values, followed by keyboard and handwriting input. Haslbeck et al. [10] also compare touch keyboard and handwriting amongst other modalities and find in addition several factors, that influence workload during driving, for example the interruptibility and the size of touch areas or handwriting input.

1.3. Research questions

To put it in a nutshell, a lot of research has been carried out on both domains: menu-driven and search-based systems. Guidelines exist about important factors, that influence the quality of each approach. Comparing these two approaches, less research could be found. Especially when focusing on cognitive workload, no study results are available that compare these different approaches to manage a menu selection in the vehicle. This research gap will be addressed in the following. The research reported here aims to investigate in more detail cognitive workload of a hierarchical menu selection and a search function with text input. The text input modes are varied between speech, touch keyboard and touch gesture (handwriting). The driving simulator study (conducted in December 2017) proves, if there are differences of these operation variations concerning cognitive workload.

2. Method

2.1. Subjects

Participants were recruited by newsletter for all employees of Porsche AG at Weissach, Germany in December 2017. In sum 36 persons participated in the study, all persons had no connection to IVIS development. Ten cases were excluded because of simulator sickness or data logging issues

The final sample consisted of 26 persons, 17 males and 9 females. One person was below 25 years old, seven participants were between 25-39, 15 between 40 and 55 years and 3 persons were beyond 55 years.

Concerning experience with the interaction methods analyzed, participants were well experienced with touchscreen and touchscreen keyboard interaction. Speech interaction was used more rarely and text input by handwriting was mostly unknown to our participants.

2.2. Apparatus

The experiment was conducted in the driving simulator of Porsche AG with motion dynamics. The mock-up was equipped with two stacked touchscreens in the center console. The lower screen was used for text input by touch gesture, other interactions were executed on the higher screen. The IVIS software prototype was especially programmed for this experiment. Touch gesture input was processed by automatic text recognition, speech input recognition was realized as a Wizard-of-Oz approach directed by a research assistant.

Gaze data (Dikablis Professional binocular eye-tracker), driving data and IVIS events were collected with a 60 Hz sampling rate and were logged within the D-Lab 3.45 software suite (time synchronized).

Fig. 1. *Impression of setup and eye-tracking measurement*

2.3. Tasks & procedure

As primary task, the participants were driving on a three-lane German highway, following a lead-vehicle. The lead-vehicle travelled with a speed varying between 65 and 75 mph. Participants were instructed to keep a constant distance between the leading car (similar driving task in Large et al. [11]).

The secondary tasks were arranged in two blocks: n-back tasks and IVIS tasks. In this form of n-back tasks, digits were presented auditory and the delayed response of the participant was carried out verbally. The higher the delay of the recall task, the higher was the cognitive workload. For further information of the n-back tasks please see Mehler et al. [12]. The n-back tasks were used to generate benchmark data to compare with the IVIS tasks. Three different levels were used: 1-back, 2-back and 3-back. For this experiment the translated version and audio files by the Chair of Ergonomics, Technical University of Munich, were used.

The IVIS tasks consisted of four different approaches to manage a menu selection: search via a menu hierarchy, search via auditory vocal text input, search via text input over keyboard on touchscreen and search via handwriting text input gesture on touchscreen. An example for a task is “Please change the interior lighting color to blue”. For exemplary procedure please see Figure2.

The menus were developed under consideration of the presented guidelines in the instruction. Menus were created with a cognitive perspective (according areas in the car) and items were sorted with respect to expected frequency of use. Five menu items were presented per page, allowing a

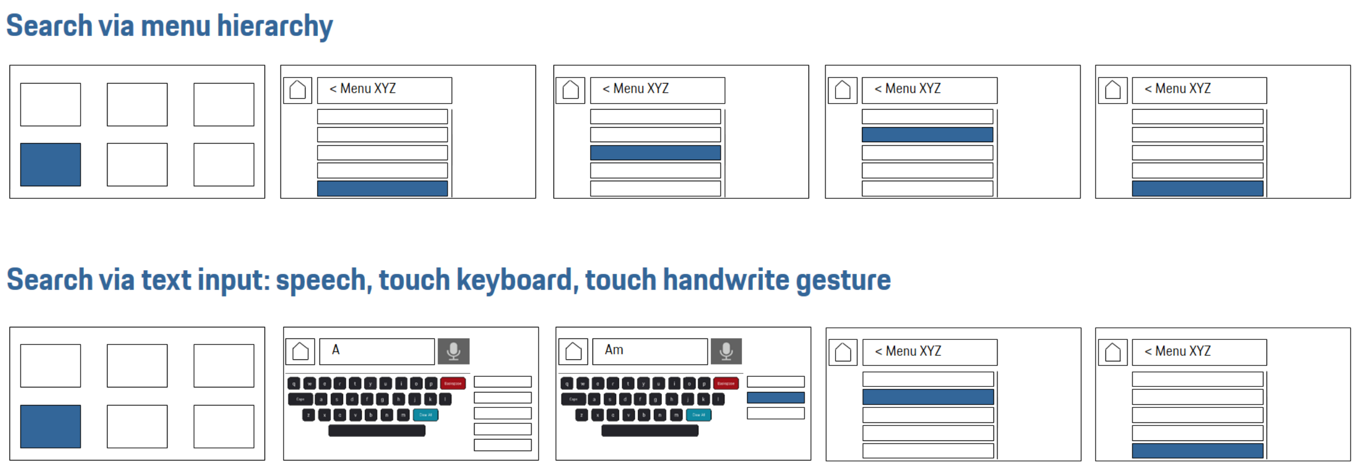

Fig. 2. *Exemplary Procedure*

reasonable touch area size. Items per menu level varied between two and sixteen entries. However, use case items were always shown on the first or second page. Regarding menu depth, the final item selection was always on the fifth and last level of the menu. Overall, menu selection use cases took 5 operating steps.

The quick search by text input use cases were constructed as follows: users selected the main menu by touch, then started entering characters (by keyboard or handwriting). After two characters were entered, the result was presented in the result list on the right side of the screen. When selected, the second to last menu page was shown and the last two items had to be selected. Overall, operating steps were comparable to those in the menu hierarchy (five steps). When entering text by speech, users had to tap the microphone button above the keyboard and then proceeded with speech input. When providing a right keyword, the research assistant forwarded the screen to the desired menu (interaction of research assistant hidden from participants). All in all, this interaction consisted also of five operating steps.

The procedure of the experiment started with a training phase to get used to the n-back and IVIS tasks. After a 5-minute test drive without secondary tasks, four blocks of secondary tasks followed: Block A with n-back tasks, Block B1 with IVIS tasks, Block B2 with a repetition of the IVIS tasks and Block B3 with a second repetition of one of the IVIS tasks. Between the subjects, Block A and B and the tasks within the Blocks were in randomized order. Between the tasks there were recovery phases without secondary tasks. The experiment had a duration of approximately 75 minutes.

2.4. Data analysis

Cognitive workload was measured by three different types of measurement: physiological data, performance metrics and subjective ratings. (O'Donnell and Eggemeier [13])

Regarding physiological data, blink-related measures (Marquardt et al. [14]) were recorded. However, due to several data-logging issues, this data is not part of the analysis. In order to measure performance within the primary task, driving data was observed. The standard deviation of distance to the lead vehicle and the standard deviation of lane position was measured (Rauch & Gradenegger [15]). Concerning secondary task performance, error-rate, number and duration of IVIS interaction events were measured. Error-rates were calculated as follows: the optimum count of operating steps was subtracted from the overall count of operating steps at this task. This balance was divided by the optimum count of operating steps to form the final error-rate. Regarding the subjective ratings, the mental dimension of the NASA TLX (Hart & Staveland [16]) was used.

To analyze differences between n-back tasks, IVIS modes and IVIS repetitions, non-parametric tests (Wilcoxon) were executed and can be found in the appendices.

In order to explore differences between interaction modes but without the effect of operating errors, a subset of error-free interactions was created. Therefore, only those tasks were considered, that had the lowest possible value of operating steps (in number 5).

3. Results

Results can be split up according to three different research questions: 1) comparing the cognitive workload of the different interaction methods to select a function; 2) examining differences in training effects of tested alternatives; 3) analysing the effect of fault tolerance and regarding faultless executions of use cases.

3.1. Cognitive workload of interaction methods

Table 1 presents the results of the n-back tasks as well as the first cycle of interaction use cases. As cognitive workload measurements, the NASA TLX mental dimension, the variability of lane position and distance keeping and error rates are reported.

NASA TLX values and error-rates seem to be quite robust indicators for the increase in cognitive workload regarding the three n-back levels, as can be seen in Table 1 and Table 3. There are significant differences between level 1 and 2 and between level 2 and 3. The variability of lane position and distance keeping on the other hand are not showing a linear increase over these three levels. 1-back and 3-back have comparable variabilities whereas 2-back shows a lower variability of these two metrics. This result should be considered when interpreting data from these measurements. Concerning the first cycle of interaction use cases, cognitive workload when searching via speech text input is

Task	N	NASA TLX [mental]		Distance	Lane position	Error-Rate	
		Mean	SD	SD	SD	Mean	SD
1-back	23	6.4	2.6	45.8	0.18	0.05	0.09
2-back	25	12.6	4.0	35.9	0.26	0.14	0.14
3-back	25	15.8	4.0	47.1	0.19	0.31	0.27
Menu 1	26	9.5	5.1	33.7	0.26	0.45	0.61
Menu 2	26	7.4	4.5	22.3	0.19	0.35	0.69
Menu 3	7	10.7	6.6	30.6	0.28	0.30	0.41
Keyboard 1	26	8.5	4.1	27.4	0.29	0.55	0.66
Keyboard 2	26	7.0	4.0	19.3	0.24	0.52	1.01
Keyboard 3	6	5.6	4.0	17.2	0.21	0.10	0.17
Gesture 1	26	8.3	4.7	38.8	0.31	1.57	1.70
Gesture 2	26	6.5	3.5	29.3	0.25	0.33	0.55
Gesture 3	7	5.0	2.7	23.4	0.25	0.00	0.00
Speech 1	26	5.6	4.2	23.2	0.24	0.01	0.20
Speech 2	26	4.5	2.6	21.6	0.20	0.00	0.13
Speech 3	6	5.7	2.6	15.2	0.14	0.06	0.13

Table 1. *Cognitive workload of interaction methods*

significantly lower than the workload while performing the other IVIS interactions. This difference is shown by the subjective measurement and the error-rates, as well as partly by the variability of the distance (differences between SDS and HWR, Menu). Regarding the remaining variants, especially input by handwriting seems to be more cognitive demanding due to its higher error rate, that also results in a higher variability of distance keeping. During the second cycle of interaction use cases, results from first cycle remain mainly constant: speech interaction is significantly less demanding regarding the subjective measurements and the error-rates. Concerning the variability of the distance, handwriting shows more variability than input by touch keyboard.

3.2. Comparing cognitive workload of n-back and IVIS

N-back tasks are useful to interpret the measurement values of cognitive workload. It is known, that 1-back represents a low to moderate cognitive workload whereas 2-back usually represents a higher workload. Compared to the n-back tasks there are following results (see Table 4).

Interaction via speech is cognitively less demanding than both N-Back levels. This is reported by subjective measurements, distance keeping variability and partly error rate. Lane keeping variability however is significant higher than at the 1-back task.

The cognitive workload of interaction with the menu hierarchy seems to be between the workload of the 1-back and 2-back levels. Concerning first round of interaction, subjective workload is higher than 1-Back and lower than 2-back. The variability of lane position is lower but the error-rate higher than at both n-back tasks. The second round of interaction has a lower workload: subjective workload is comparable to 1-back and lower to 2-back, variability of distance keeping is lower, variability is lower than 1-back and comparable to 2-back and error-rate is higher than 1-back and comparable to 2-back.

The cognitive workload of keyboard text input is more hardly to interpret, because results between the measurement methods are not homogeneous. First round of interaction is subjective more demanding than 1-back and less demanding than 2-Back. Variability of distance is lower than both n-back levels, variability of lane position and error rates are higher than both levels. The workload of the repetition is comparable to 1-back and less demanding than 2-back. Variability of distance keeping is lower than both levels and variability of lane position and error rates are comparable to 2-back.

Regarding workload of handwrite recognition, the first round of interaction, subjectively workload is comparable to 1-back, variability of distance keeping is comparable to both levels and variability of lane position and error rates are higher than both levels. The second round of interaction with handwrite interaction shows similar results. Only variability of lane position is now comparable to 2-back and error rates are comparable to both levels.

Cognitive workload of all IVIS interactions seems to be below the workload of 3-back tasks. This is shown by subjective measurements and variability of distance.

3.3. Training effects concerning cognitive workload

Interactions with the IVIS were repeated for two times, the cognitive workload measurements for these trials are shown in Table 1 and Table 5. Between first and second trial, cognitive workload decreases especially at the menu interaction and the handwriting task. Regarding the menu task, there are significant decreases at the subjective measurement, as well as the variability of distance and lane position. Concerning handwriting, the differences of the subjective measurement and the error rates are significant lower. Regarding the touch keyboard task, there are only significant decreases at the lane position variability. Speech interaction on the other hand shows no significant differences and remains mostly on the same level.

Regarding trials two and three, cognitive workload seems to rise at some tasks, but these differences are not significant (could be due to small sample size in third trial). There is only one significant drop in subjective cognitive workload within the keyboard task repetitions.

3.4. Comparing error-free trials

Text-input by speech was mostly error-free due to its wizard-of-Oz approach. In order to focus on the differences on the way of interaction and not on the error-rate, results presented here, are only focusing on error-free trials (Table

2). Due to the unfamiliarity with the system, numerous errors occurred especially in the first trial. Therefore, results are presented for the second trial with a higher sample size and statistical significant differences (Table 6): speech is less demanding than menu and the keyboard task.

Task	N	NASA TLX [mental]	Distance	Lane position
		Mean	SD	SD
1-back	16	6.1	49.4	0.20
2-back	6	12.1	35.1	0.11
3-back	0	-	-	-
M2	11	6.8	15.8	0.15
K2	8	6.4	15.3	0.20
G2	7	5.0	22.9	0.16
S2	19	4.7	20.9	0.20

Table 2. *Cognitive workload of error-free second trials (M=Menu, K=Keyboard, G=Gesture, S=Speech)*

Concerning variability of distance keeping there are no significant differences, concerning variability of lane position there are significant differences between second trial of menu and speech tasks.

4. Discussion and Conclusions

The study examines the cognitive workload of several methods to manage a function selection with an IVIS: selection via menu hierarchy or selection via search and text input. Text input methods are varied between speech, touch keyboard and touch gesture handwriting. Results show, that cognitive workload of the search function with text input mode via speech is lower than of the remaining variants. Strayer et al. [3] also found a difference between voice interaction and interaction via center console, which supports this finding.

There seem to be no differences in cognitive workload between the haptic interactions presented in this experiment, although conceptual differences seem to be quite large. Only handwriting input is partly more demanding, especially due to its significant higher error rate.

Comparing workload of n-back tasks and IVIS tasks, speech interaction is less demanding than both n-back levels. Interaction via menu hierarchy has a workload between 1-back and 2-back. A comparison of handwriting input and keyboard input with n-back levels is difficult, because measurement methods are varying strongly. When looking at subjective measurements, keyboard input as well as handwriting input is comparable to the workload of 1-back.

Regarding training effects of IVIS interactions, cognitive workload of menu interactions and handwriting interactions are decreasing significantly. Maybe hierarchical menus need some training to know more about the logic of the structure and touch gesture inputs need some training how characters can be recognized by the system. The implemented speech task was quite fault-tolerant because of its Wizard-of-Oz approach. Finally, text input by touch keyboard is often used by smartphone users, which could explain, that there are not so many training effects concerning these variants.

An additional analysis of error-free task trials shows, that the impact of operating errors on the cognitive workload should be considered. Workload between the alternatives is more aligned in comparison to trials that includes errors. When discussing this topic, it should be kept in mind, that fault tolerance also is often a characteristic of an interaction alternative. The freedom of design when creating menus is more dynamic and leaves a higher risk of decreasing fault tolerance than solely technical implementations of text input methods.

Nevertheless, the impact of operating errors on the cognitive workload should be addressed in further studies. An example would be a fault-tolerance-factor compared to the well-known age-factor in key-stroke modellings.

5. References

[1] Lee, E. S., & Raymond, D. R. (1993). Menu-driven systems. *Encyclopedia of Microcomputers*, *11*, 101-127.

[2] Heinrich, C. (2013). Fighting driver distraction: worldwide approaches. In Proceedings of the 23rd International Technical Conference on the Enhanced Safety of Vehicles (ESV).

[3] Strayer, D., Cooper, J., Goethe, R., McCarty, M., Getty, D., Biondi, F. (2017). Visual and Cognitive Demands of Using In-Vehicle Infotainment Systems. Salt Lake City, UT: University of Utah.

[4] Norman, K. L. (2008). Better design of menu selection systems through cognitive psychology and human factors. Human factors, 50(3), 556-559.

[5] Burnett, G. E., Lawson, G., Donkor, R., & Kuriyagawa, Y. (2013). Menu hierarchies for in-vehicle user-interfaces: Modelling the depth vs. breadth trade-off. Displays, 34(4), 241-249.

[6] Hornof, A. J., & Halverson, T. (2003, April). Cognitive strategies and eye movements for searching hierarchical computer displays. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 249-256). ACM.

[7] Takashi, M. (2008). Investigating the cognitive loads involved in searching hierarchical menus using eyeblinking as an index. The Japanese Journal of Cognitive Psychology, 6(1) 1-10.

[8] Graf, S., Spiessl, W., Schmidt, A., Winter, A., & Rigoll, G. (2008, April). In-car interaction using search-based user interfaces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1685-1688). ACM.

[9] Kujala, T., & Grahn, H. (2017, September). Visual Distraction Effects of In-Car Text Entry Methods: Comparing Keyboard, Handwriting and Voice Recognition. In Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 1-10). ACM.

[10] Haslbeck, A., Popova, S., Krause, M., Pecot, K., Mayer, J., & Bengler, K. (2011, July). Experimental evaluations of touch interaction considering automotive requirements. In International Conference on Human-Computer Interaction (pp. 23-32). Springer, Berlin, Heidelberg.

[11] Large, D. R., Burnett, G., Anyasodo, B., & Skrypchuk, L. (2016, October). Assessing Cognitive Demand during Natural Language Interactions with a Digital Driving Assistant. In Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 67-74). ACM.)

[12] Mehler, B., Reimer, B., & Dusek, J. A. (2011). MIT AgeLab delayed digit recall task (n-back). Cambridge, MA: Massachusetts Institute of Technology.

[13] O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology.

[14] Marquart, G., Cabrall, C., & de Winter, J. (2015). Review of eye-related measures of drivers' mental workload. Procedia Manufacturing, 3, 2854-2861.

[15] Rauch, N., & Gradenegger, B. (2007). Das Konzept des Situationsbewusstseins und seine Implikationen für die Fahrsicherheit. FAT-Schriftenreihe, (210).

[16] Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Advances in psychology, 52, 139-183.

6. Appendices

Table 3. *Results of wilcoxon tests comparing cognitive workload of interaction methods (M=Menu, K=Keyboard, G=Gesture, S=Speech)*

Task	NASA TLX [mental]	SD Distance	SD Lane position	Error-Rate
n-back				
1 - 2	.000	.784	.248	.004
2 - 3	.000	.391	.214	.005
IVIS 1				
M – K	.354	.469	.517	.537
M – G	.646	.166	.353	.028
M – S	.002	.038	.292	.003
K – G	.852	.058	.648	.044
K – S	.010	.367	.080	.001
G – S	.003	.001	.269	.000
IVIS 2				
M – K	.722	.585	.166	.474
M – G	.569	.115	.657	.948
M – S	.000	.829	.778	.023
K – G	.852	.030	.957	.447
K – S	.000	.620	.191	.001
G – S	.000	.264	.326	.004

Table 4. *Results of wilcoxon tests comparing cognitive workload of interactions methods and N-Back tasks (M=Menu, K=Keyboard, G=Gesture, S=Speech)*

Task	NASA TLX [mental]	SD Distance	SD Lane position	Error-Rate
1-back vs. IVIS 1				
M	.009	.130	.000	.002
K	.025	.018	.000	.001
S	.124	.001	.003	.396
G	.137	.304	.001	.002
1-back vs. IVIS 2				
M	.533	.002	.006	.050
K	.737	.000	.001	.007
S	.015	.003	.002	.108
G	.879	.114	.003	.089
2-back vs. IVIS 1				
M	.011	.276	.032	.011
K	.004	.006	.023	.005
S	.000	.001	.069	.028
G	.002	.657	.020	.002
2-back vs. IVIS 2				
M	.000	.007	.192	.520
K	.000	.000	.074	.126
S	.000	.003	.174	.349
G	.000	.241	.162	.005
3-back vs. IVIS 1				
M	.000	.150	.000	.338
K	.000	.002	.000	.306
S	.000	.002	.002	.000
G	.000	.600	.000	.007
3-back vs. IVIS 2				
M	.000	.002	.004	.436
K	.000	.000	.001	.475
S	.000	.002	.002	.531
G	.000	.022	.003	.000

Table 5. *Results of wilcoxon tests comparing cognitive workload of IVIS repetitions (M=Menu, K=Keyboard, G=Gesture, S=Speech)*

Task	NASA TLX [mental]	SD Distance	SD Lane position	Error-Rate
IVIS 1 vs. 2				
M	.012	.013	.009	.145
K	.062	.073	.034	.361
S	.107	.551	.292	.763
G	.038	.242	.074	.006
IVIS 2 vs. 3				
M	.715	.735	.397	.109
K	.027	.345	.600	.285
S	1.000	.753	.463	.317
G	.068	.686	.686	.655

Table 6. *Results of wilcoxon tests comparing cognitive workload of IVIS second trial, error-free (M=Menu, K=Keyboard, G=Gesture, S=Speech)*

Task	NASA TLX [mental]	SD Distance	SD Lane position
IVIS 2			
M – K	.180	.686	.500
M – G	.180	.655	.180
M – S	.042	.953	.028
K – G	.655	.655	.655
K – S	.016	.779	.401
G – S	.197	.753	.515

Digitalisation in the infotainment: User needs and requirements – an explorative approach.

Anna Pätzold*[1], Rami Zarife[2], Michael Wagner[2], and Josef F. Krems[3]

[1, 2] EE Core Infotainment & Program Management, Opel Automobile GmbH, Rüsselsheim, Germany
[3] Cognitive and Engineering Psychology, Institute for Psychology, Technische Universität Chemnitz, Germany
*anna.paetzold@opel.com

Abstract:
In order to understand drivers' needs and requirements in extending infotainment functions, an explorative approach, consisting of creativity workshops, a focus group and an online survey was pursued. In the creativity workshops and the focus group, spending the driving time usefully was identified as the main motivational factor for drivers to engage into their mobile devices while driving. Nonetheless, they did not want to be distracted. The need to be informed about the environment, including participants' social network and traffic circumstances, was highlighted. The online survey found interaction effects between modality of secondary task and driving situation. Context factors were found to have different effects on the willingness to engage in the secondary task in question. Especially for the context factor *street type*, the demanded secondary task modality effect showed the highest impact. The cascade of the explorative approach provided a feasible way to obtain a comprehensive understanding of driver needs and requirements in extending infotainment functions.

1. Introduction

Due to the digital revolution, new and extended functions will be available both on smartphones and in the in-car infotainment systems [1, 2], increasing the amount of information provided to the driver [3].

As a visually-manually focused task [ibid.], driving interferes with any other task demanding the same modalities [4]. According to the Task-Capability-Interface-Model, an imbalance between a driver's capabilities and the task demands can lead to a loss of control [5].

1.1. Engagement in secondary tasks while driving

Although negative effects of engagement into secondary tasks on reaction times [6], visual monitoring [7] and vehicle control [8, 9], including speed and lane keeping, were found, and the usage of smartphones while driving is banned in many countries, drivers today use their mobile phones and personal digital assistants more frequently while driving [10, 11, 12].

The recent US-American naturalistic driving study SHRP2 found an increase in crash risk due to operating in-vehicle devices by an odds ratio of 2.5, leading to 3.53 % of all observed accidents [13]. Further, the usage of nomadic devices while driving was found to have an odds ratio of 3.6 causing 6.40 % of all observed accidents [ibid.]. SHRP2 also found distracting activities, such as smartphone usage, to occur much more frequently than drivers' impairments, such as drowsiness [14].

Equally, the European naturalistic driving study UDRIVE found the most distracting activities to be primarily located in the middle console [15]. From all the observed secondary tasks, mobile phone usage was the most frequent executed and had the longest task engagement duration [16].

1.2. Motivations for engaging in secondary tasks

As identified in a review [17], the main key themes for engagement in distracting activities in distraction research are perceived risk and incidence of use. Though, parameters influencing perceived risk are still missing.

As the drivers' needs change depending on the context [18, 19, 20, 21], one motivation for engaging in secondary tasks while driving can be the context. Further, the need for information on the environment, such as traffic and communication were found as influencing factors [22].

1.3. Aim and scope of the current research

In order to understand driver's needs and requirements in extending infotainment functions, an explorative approach, consisting of creativity workshops, a focus group and an online survey, was pursued (Fig. 1).

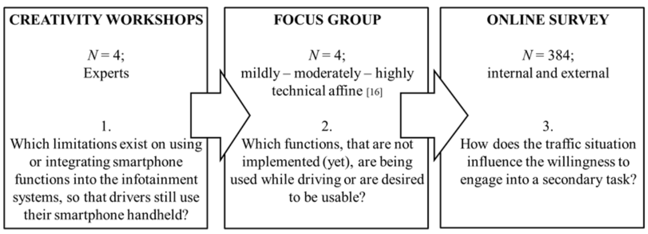

Fig. 1. *Methodology of the explorative approach*

2. Creativity Workshops

2.1. Research Questions

Phone projection applications, such as AndroidAuto and Apple CarPlay, give the possibility to use specific smartphone functions while driving, intending to make the handheld use of the smartphone while driving superfluous.

Still, some smartphone functions are not yet implementable into the infotainment systems or are not suitable for usage while driving. In order to investigate these factors, the limitations of smartphone functions as well as potential HMI characteristics were explored.

2.2. Method

Two creativity workshops were conducted with each $N = 4$ internal experts in infotainment HMI engineering. The first workshop used the Double Reverse Technique [23, Fig. 2], and was intended to identify elements of smartphone functions that make these functions uncomfortable to use or restrict them from using while driving. Smartphone functions were categorised into communication, navigation, media, browsing and other (Table 1).

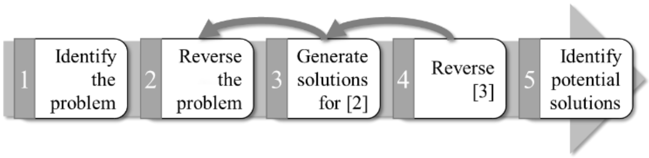

Fig. 2. *Double Reverse Technique [23]*

Table 1. Smartphone functions.

	Examples
Communication	Messenger, Calls, Address Book, Social Media,…
Navigation	Traffic information, POIs, Favorites,…
Media	Streaming Services, Playlists, Gallery,…
Browsing	Search, Shopping, Finances,…
Other	Clock, Calendar, Notes,…

The second workshop used the Brute Think Technique [ibid.,Fig. 3] to identify HMI characteristics that can be used to implement these solutions.

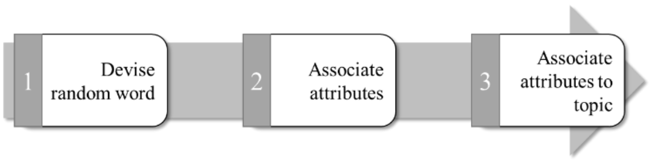

Fig. 3. *Brute Think Technique [23]*

2.3. Results

Selected results of both workshops are shown in Table 2.

Table 2. Results Creativity Workshops.

	Problem	Double Reverse Result	Brute Think Result
Communication: Messenger	Long message being read-out or displayed	Interruptible reading-out	Highlighting of relevant information
Navigation: Points-of-Interest	Irrelevant POIs	Status-based selection: empty gas tank, gas stations at top of the list	Selection of nearby POIs based on user behaviour
Media: Playlists	Many playlists to select from		Tiles for each playlist
Browsing: Search	High input quantity	Use speech	Minimize data entry
Other: Calendar	Not synced	Adapt Address Book, Navigation, …	

2.3.1 Smartphone Functions. For the in-car use while driving, too much information is shown. In addition, many input steps are necessary to execute the intended function. Using the smartphone while driving is uncomfortable; not only because of hand position, the position of the centre-stack display, or the provoked distraction, but also because of the cognitive dissonance perceived by drivers. Since drivers are aware of the distracting effects of smartphone usage, they experience a conflict between their need to engage in the smartphone and their need to avoid distraction while driving.

2.3.2 HMI Characteristics. In order to address these issues, the HMI can be changed by integrating new elements e.g. highlighting information, introducing shortcuts to recently, frequently or intended to-be-used functions, change the modality of the input and output. Further, position of the shown information can be adapted between and within displays.

2.4. Conclusion

Since the overall issue is the amount of displayed information, the infotainment system should be able to provide the same content with less characters. Since the driving task is visual-manual focussed, the secondary task modality should potentially load on another modality, such as cognitive-auditory.

3. Focus Group

3.1. Research Questions

A focus group [24] was conducted to further investigate driver's motivation to use a smartphone while driving. It was of interest (1) which mobile devices and functions participants currently use while driving, (2) which functions they would like to be able to use in their cars besides mobile devices' functions, (3) which strategies they use to avoid distraction and (4) potential design solutions to improve usage.

3.2. Method

N = 4 participants were chosen out of the company's internal participants pool based on their technical affinity. Technical Affinity was assessed beforehand via an online screener using the Questionnaire on Technical Affinity (TA-EG [25]). According to the distribution, one participant of the 33rd, one of the 66th, and two of the upper percentile participated. Three male and one female participants took part, with a mean age of *M* = 43.5 years (*SD* = 13.08, Range = 26-54 years).

The first part consisted of participants individually filling a worksheet asking for currently in-car used nomadic devices, desired functions, strategies to avoid distraction and potential designs to improve usage. The second part consisted of an open discussion, debating an order and requirements for preferred implemented features.

The focus group was recorded on video. Participants agreed on video recording before filling the online survey. The recorded video was transcribed using ELAN 4.9.4 and FreeQDA.

3.3. Results

3.3.1 Currently in-car used nomadic devices. Besides the less technical affine participant, participants stressed the wish to use the smartphone while driving to communicate and to navigate, especially when their in-car navigation systems did not provide live traffic.

Communication included phoning via Bluetooth (*n* = 4), dialling via speech recognition (*n* = 2), reading and typing messages (*n* = 2), speech-based texting (*n* = 2).

Smartphone-based navigation was used by three participants and by two of them via phone projection applications, such as AndroidAuto and Apple CarPlay. It was also used on familiar routes to be informed about live traffic.

One participant also used Music Streaming on a daily basis via AndroidAuto.

3.3.2 Desired functions. Speech recognition systems are currently used and desired to provide natural language understanding (*n* = 3). Further, one participants wished to be able to listen to and record voice messages while driving.

Also, a synchronisation between personal mobile devices and the infotainment system regarding data and files was mentioned (*n* = 1). As a part of their daily routine, the car shall be able to act as an office provider.

Further, participants wished for a more stable internet connection via W-LAN in their cars (*n* = 2).

3.3.3 Strategies to avoid distraction. The technically less affine participant mentioned to avoid controlling any infotainment function while driving. Both the technically less and moderate participants stated to put their smartphones out of reach while driving, and only using it in traffic jams (*n* = 1).

In order to monitor, the technically moderate and high affine participants mentioned to switch their gazes more frequently between infotainment displays and the traffic scene.

Phone projection applications were mentioned to avoid distraction (*n* = 2), especially when used with speech recognition (*n* = 1). Smartphone-based functions like navigation and playlists were set up before starting to drive (*n* =2).

If the smartphone was used while driving, it was held in the right hand next to the steering wheel or the hand was laying on the right thigh.

3.3.4 Potential designs to improve usage. Three participants were experienced with Head-up displays and mentioned the advantages, as they did not need to take their gazes far from the traffic scene. Regarding input devices, participants were indecisive on rotary push, touchscreens, touchpads and steering wheel controls. The ability to control them blindly was highlighted. They agreed on the importance of a haptic feedback and a system that requires few input steps, by i.e. providing suggestions.

Further, the technically high affine participants mentioned new input technologies such as eye tracking, to be an interesting and compelling approach.

Participants reached consensus on the need for a minimized distractive system that still fulfils their needs. Therefore, the usage of the infotainment system shall be easy and intuitively understandable. That is, interaction methods shall be indicated clearly and unambiguously. Easiness and efficiency were stated to be most important. As an example, one participant said *"there are several easy things, I press a button instead of telling the system to warm up the ventilation"*. Another one reported problems with speech recognition, as he has *"never yelled at a system that often before"*.

Further, they wished to have adapting or customizable display and control elements with their most frequently used function. Especially when a car is shared, an automatic adaptation of infotainment and vehicle parameters, such as seating, was mentioned (*n* = 3).

3.3.5 Open discussion. The two technically less and moderate affine participants mentioned avoidance of smartphone use while driving, since their cars' infotainment systems does not have phone projection applications. The other two, technically high affine participants, use AndroidAuto or Apple CarPlay daily, but still missed some functionalities. Therefore, they intentionally disconnect their smartphones due to restricted functions, i.e. scrolling down long lists, or not implemented functions, i.e. recording voice messages.

Further, one participant mentioned to *"use the smartphone to receive, read and write text messages, which is not optimal in every driving situation"*. Participants agreed that one main factor for the decision whether or not to use their smartphone while driving was the driving situation. One participant stated, that he rather engages into a secondary tasks when he can foresee the upcoming situation. Driving on a highway with moderate traffic seems more anticipative to him than driving in a city scenario.

3.4. Conclusion

It was especially difficult for participants to think of further functions they would like to use while driving. Although no needs for future functions could be retained from the focus group, it gave a good impression on functionalities that are designed unsuitably or even irritating.

Supporting [22] findings, the need to be informed about the environment, including participant's social network and traffic circumstances, was empathised and stated to contradict with the need to not be distracted.

The need to be informed about the social and traffic environment were the main motivational factors for using the smartphone while driving.

4. Online Survey

4.1. Motivation and Research Questions

Although interfering effects in dual-task execution can be explained using the Multiple Resource Theory [4], little research on the interaction of driving situation variables and secondary task execution was done.

In a survey study, Ferreira et al. [53] identified drivers to be least likely to engage in their phones on city roads, but rather on highways. Young and Lenné [55] found in an online survey, that secondary tasks while driving were avoided in bad weather, winding roads, heavy traffic or night. Supporting these findings, Britschgi et al. [52] identified bad weather, heavy traffic and city roads to influence the willingness to use a phone while driving. Hancox et al. [54] found drivers decision (not) to engage in a phone task, such as placing or answering calls and sending or reading texts, to be depended both on the perceived demands of the roadway and the phone function. Especially placing or answering a phone call was of low willingness in high demanding driving situations.

Regarding driving situation complexity, Fastenmeier [27] found street characteristics to have the highest impact on driving situation complexity, whereas traffic density and visibility were identified as weighing factors.

Horberry et al. [26] found complex driving situations to lead to compensatory behaviour and higher perceived distraction when simultaneously executing an in-vehicle entertainment task or talking on the smartphone. According to Lerner et al. [28], task-related motivations to be dominant decision factors in contrast to driving-related motivations, such as the upcoming driving maneuver.

In UDRIVE it was also found, that the willingness to engage in a secondary task depended on the workload of the task [16]. Contrary to the hypotheses, complex tasks were more likely executed in complex driving tasks and also longer in duration.

As the focus group revealed, the decision on whether or not to engage into a secondary task while driving seems to be depending on the driving situation and the modality of the task, hence, the interaction between the two was the focus for the online survey.

4.2. Method

In order to investigate the effect of the driving situation on the willingness to engage in a secondary task, an online survey was conducted. An online survey was chosen to avoid social desirability by providing anonymity [e.g. 29].

4.2.1 Participants. All participants held a valid driver's license. N = 444 persons participated in the online survey, whereas n = 60 had to be removed due to incomplete data. Participants (23.7 % female) were M = 45.08 years old (SD = 9.57, Range: 20-75 years). They were recruited via the company's internal participants pool, university's student mailing lists and social media platforms. As an incentive, two 25€ Amazon vouchers were drawn among interested non-company participants.

4.2.2 Measures. On demographics, age, gender, driver license possession and year of acquisition, annual mileage and eye vision were asked.

Technical affinity was assessed using the TA-EG [25]. Driving Style was rated on the short version of the Multidimensional Driving Style Inventory [30], adapted to and validated in Europe by [31]. Two items on the wish to use and connect the smartphone with the infotainment system were included [21]. The knowledge on and usage of new media were assessed. Further, the willingness to engage in a secondary task depending on the driving situation was investigated using a choice-based conjoint analysis (CBCA).

4.2.3 Context. Driving situation profiles were generated using the context factors adapted from [18, 21]. A driving situation was defined by street type (city, rural, highway), landscape (flat, hills, trees), traffic density (low, moderate, high), weather (dry, rain, snow) and daytime (day, night) (Fig. 4). See Table 8 in Appendices.

In order to reduce the number of profiles, two orthogonal arrays [32] were combined. For three-step factors, a fractional 3^4 Design [33] or Plan 3 [32] was applied (Table 8), whereas daytime was applied using the first nine columns and first 18 rows of Plan 8 [ibid.]. Profiles were assorted to 18 choice sets using a Balanced Incomplete Design [33, 34]. Participants were randomly assigned to three groups, each being presented six choice sets.

Since spontaneous answers in a low-involvement situation [33, 35] and a low social desirability [36] were required, a CBCA was chosen to assess willingness to engage in a secondary task in a specific driving situation. Participants were asked to choose the one driving situation in which they would not engage in the secondary task. Alternatively, they could choose the none-option of *"I would use the function in all scenarios"*. Whether the task was to be executed on an in-vehicle display or a hand-held device was not of importance. See Fig. 5 for an example of the online survey.

Fig. 4. *Driving situations used in the online survey*

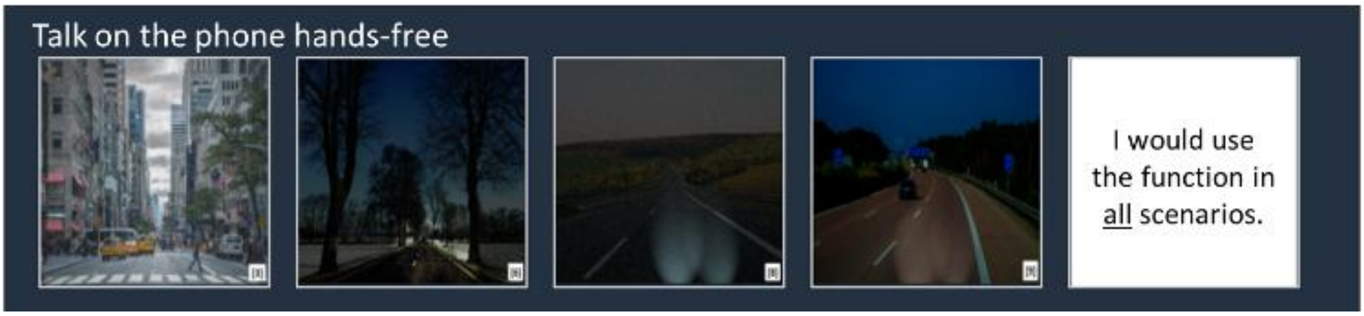

Fig. 5. *Example of the CBCA from the online survey*

4.2.4 Secondary Tasks. Following the Multiple Resource Theory [4] secondary tasks covering the four modalities, both encoding strategies and interaction styles [37] were evaluated. See Table 3 for the six secondary tasks.

Table 3. Secondary Tasks following [4, 37]

Task	Modality	Encoding	Interaction
read a text message	visual	verbal	passive
type a text message	visual-manual	verbal-spatial	active
watch a video	visual-auditory	verbal	passive
talk on the phone hands-free	cognitive-auditory	verbal	active-passive
make a shopping list	cognitive	verbal	active
adjust volume	manual	spatial	active

4.3. Results

4.3.1 Connectivity. The results for the willingness to use content of electronic devices and the wish to connect the smartphone with the in-car infotainment system [21] are shown in Fig. 6 and Fig. 7. Most of the participants want to only consume content without necessarily sharing their experiences and content. Also, most participants want to be able to connect their smartphones with the infotainment system.

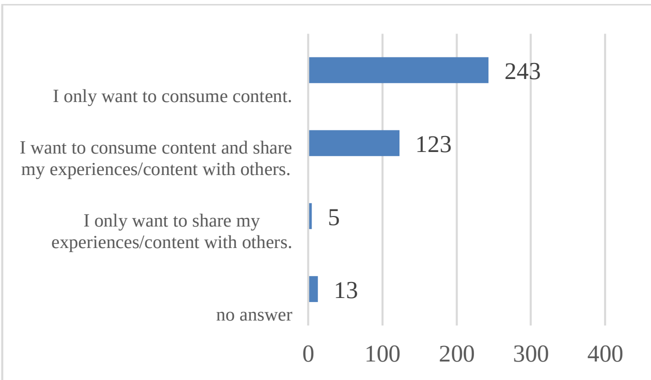

Fig. 6. *Willingness to use electronic devices, N = 384*

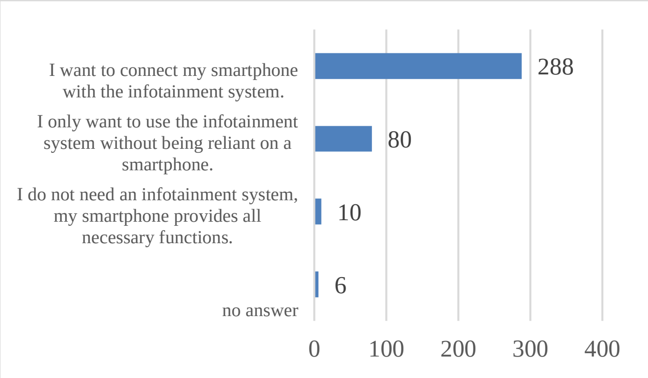

Fig. 7. *Willingness to connect the smartphone and infotainment system, N = 384*

4.3.2 Technical Affinity. In order to control for a normal distribution of technically less, moderately and highly affine participants, percentiles of the TA-EG [25] scores were calculated. For further analysis, technical affinity was split into low (TA-EG score: 9.25-12.55, n = 124), moderate (12.6-13.65, n = 123) and high (13.65-17.2, n = 137).

4.3.3 Driver Profiles. In order to identify driver types, item scores in the short MDSI were multiplied with the adjusted factor loadings, generating factor scores. See Table 4 for results. Driver profiles were extracted by normalizing the factor scores [31], Fig. 8. Ratings of n = 70 participants did not exceed the threshold for one driver profile category. In total, 59.55 % of participants ratings did not load on more than one factor of the MDSI, hence, further analysis only took the six driving styles of angry (ANG), risky (RIS), anxious (ANX), dissociative (DIS), careful (CAR) and distress-reducing (DRE) driving into account.

Table 4. Driver profiles, N = 384

	ANG	RIS	ANX	DIS	CAR	DRE
n	76	65	70	55	62	56

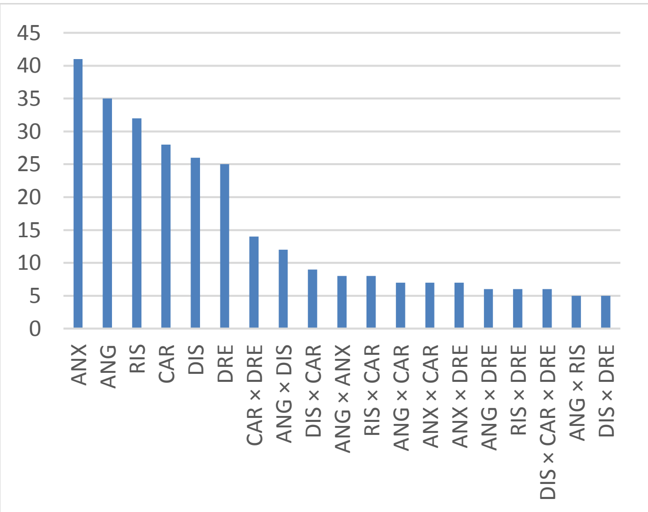

Fig. 8. *Distribution of Driver Profiles MDSI [17], n = 287*

Note: Only Driver Profiles with $n \geq 5$ are shown.

4.3.4 Engagement in secondary tasks. Fig. 9 shows the results of the CBCA on willingness to engage in the secondary task while driving for the driving situation factors. A higher percentage indicates a higher probability of a decision against engaging in the secondary task.

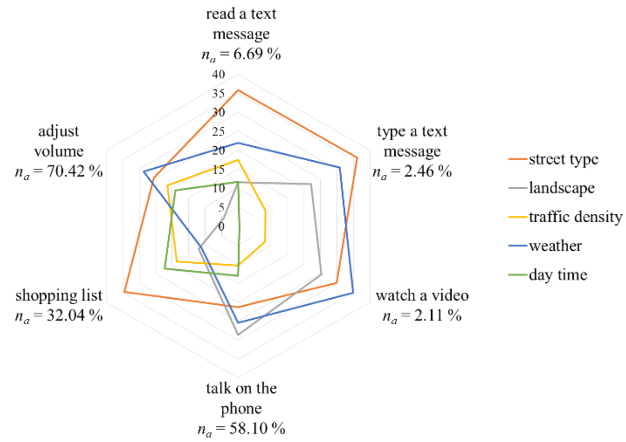

Fig. 9. *Relative importance values of context factors for each secondary task*

Note: n_a describes the percentage of participants willing to use the function in every driving situation.

Relative importance values of the context factors and levels found in the CBCA for the decision against engagement for each secondary task are shown in the following figures 10 - 15. Please find the path-worth utilities in

Table 910, Appendices.

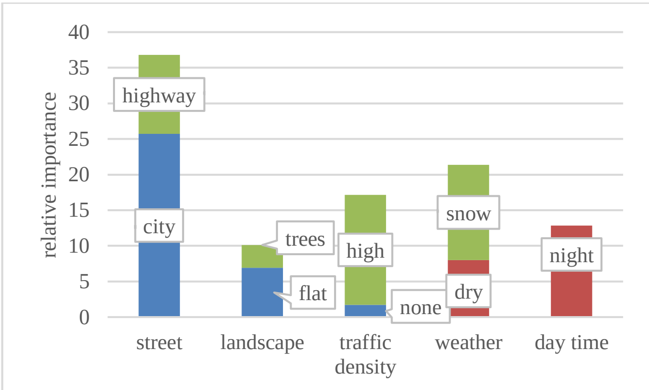

Fig. 10. *Read a text message*

Note: Relative importance for context factors: street 36.79%, landscape 10.12 %, traffic density 17.16%, weather 21.38%, day time 12.84%

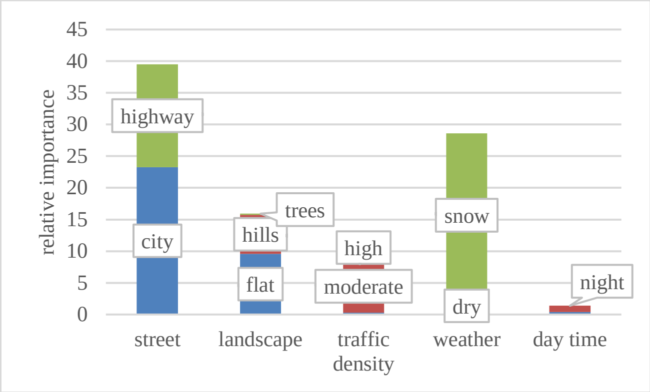

Fig. 11. *Type a text message*

Note: Relative importance for context factors: street 39.52%, landscape 15.93 %, traffic density 12.53%, weather 28.61%, day time 1.38%

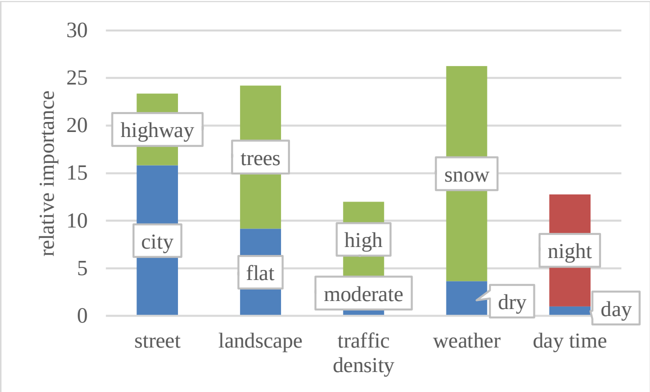

Fig. 12. *Watch a video*

Note: Relative importance for context factors: street 23.37%, landscape 24.21 %, traffic density 11.98%, weather 26.24%, day time 12.77%

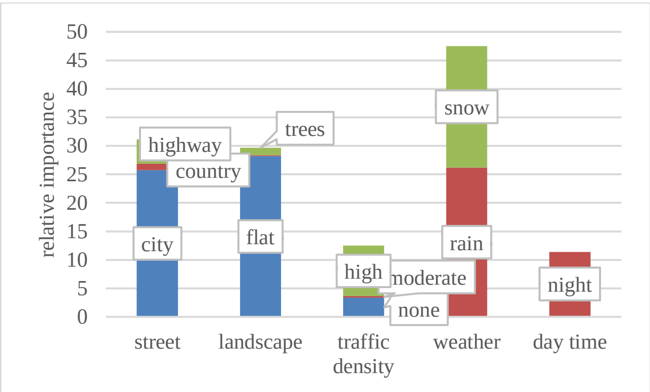

Fig. 13. *Talk on the phone hands-free*

Note: Relative importance for context factors: street 31.11%, landscape 29.66%, traffic density 12.49%, weather 47.52%, day time 11.38%

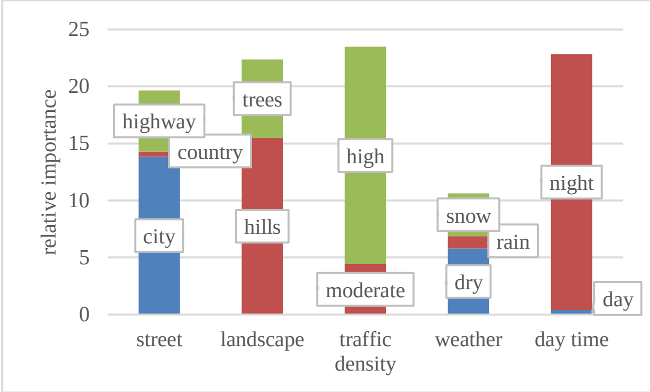

Fig. 14. *Make a shopping list mentally*

Note: Relative importance for context factors: street 19.65%, landscape 22.37%, traffic density 23.48%, weather 10.62%, day time 22.86%

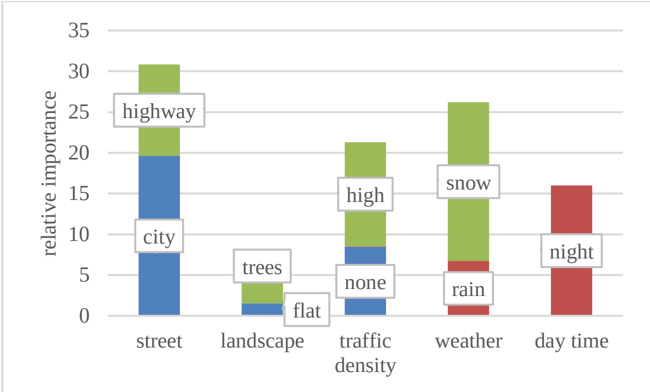

Fig. 15. *Adjust volume manually*

Note: Relative importance for context factors: street 30.82%, landscape 4.78%, traffic density 21.29%, weather 26.19%, day time 15.98%

4.3.5 Clusters. To investigate influence factors on the decision not to engage in a secondary task, hierarchical cluster analyses were calculated for each secondary task [36, 38]. For all secondary tasks, the dendrogram identified two clusters. Cluster A included participants deciding against engaging in the secondary task depending on the driving situation, cluster B included participants willing to engage in the secondary task in every driving situation. Table 5 shows the cluster groups for each secondary task. Driver profiles associated with the clusters are shown in Fig. 16.

Table 5. Clusters for secondary tasks, N = 384

	$n_{Cluster\ A}$	$n_{Cluster\ B}$
read a text message	361, 24.1% female	23, 17.4% female
type a text message	374, 23.5% female	10, 20.0% female
watch a video	376, 23.1% female	8, 37.5% female
talk on the phone hands-free	217, 24.4% female	167, 22.2% female
make a shopping list mentally	120, 26.7% female	264, 22.3% female
adjust volume manually	266, 22.6% female	118, 25.4% female

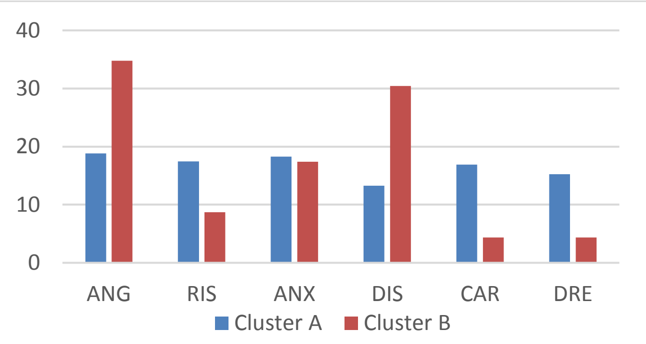

Fig. 16. *Percentages of Driver Profiles for Cluster A and Cluster B, N = 384*

Table 7 shows the effects of the cluster characteristics on the willingness to engage in the secondary task. In the following, significant effects for each secondary task are described in detail.

Read a text message. Participants with a mean age of 40.74 years (*SD* = 8.49) were more willing to engage in the secondary task in every driving situation than participants of *M* = 45.36 years (*SD* = 9.57). Participants willing to read a text message in every driving situation were rather classified as more angry and dissociative drivers and less careful and distress-reducing drivers.

Watch a video. Participants with a higher annual mileage (*M* = 26 250.00 km, *SD* = 9 543.14) were more willing to engage in the secondary task than participants with a lower annual mileage (*M* = 17 816.91 km, *SD* = 11 191.67).

Talk on the phone hands-free. Participants of *M* = 44.07 years (*SD* = 9.51) were more willing to make a hands-free phone call while driving than participants with a mean age of 46.29 years (*SD* = 9.11). Participants with a higher annual mileage (*M* = 19 407.69 km, *SD* = 12 246.64) were more willing to engage in the secondary task than participants with a lower annual mileage (*M* = 16 585.12 km, *SD* = 9 489.63).

Make a shopping list mentally. The two clusters differed significantly regarding age, as participants of *M* = 43.58 years (*SD* = 10.218) were more willing to engage in the secondary task than participants of a mean age of 45.76 years, *SD* = 9.19).

Adjust the volume manually. Participants willing to engage into the secondary tasks were 44.43 years (*SD* = 9.64) on average, whereas participants deciding against the secondary tasks were *M* = 46.65 years old (*SD* = 9.07). Participants with a higher annual mileage (*M* = 18 811.90 km, *SD* = 11 595.73) would rather engage in adjusting the music volume in every driving situation than those with a lower (*M* = 16 145.70 km, *SD* = 10 106.88, *F* = 1.096, *p* = .296, η_p^2 = .003).

Table 6. Effects of the identified clusters on the willingness

		Read a text message	Type a text message	Watch a video	Make a phone call hands-free	Make a shopping list mentally	Adjust the music volume
Age	*p*	.025*	.098	.203	.021*	.038*	.035*
	η_p^2	.013	.007	*.004*	.014	.011	.012
Gender	*p*	.463	.795	.343	.603	.356	.541
	CC	.037	.013	.048	.027	.047	.031
Annual	*p*	.296	.277	.035*	.014*	.180	.296
Mileage	η_p^2	.003	.003	.012	.016	.005	.003
Technical	*p*	.201	.853	.599	.110	.616	.461
Affinity	η_p^2	.004	.000	.001	.007	.001	.001
Driver	*p*	.034*	.442	.432	.554	.314	.084
Profile	*CC*	.177	.112	.113	.102	.124	.159

4.3.6 Calculations for driving situations. Over all secondary tasks, the driving situation that raised the highest willingness to engage into any secondary task, independent of its modality, is country road (0.02), flat (0.43), moderate traffic (0.26), and dry weather (0.28) by day time (0.03) with a total utility of 1.02.

Situations that raised the lowest and the highest willingness to engage in the secondary task in question are shown in Table 7.

Table 7. Calculations for willingness to engage in a secondary task

		Street Type	Landscape	Traffic Density	Weather	Day Time	Total utility
read a text message	lowest	Country 0.00	Hills 0.00	Moderate 0.00	Dry 0.00	Day 0.00	0.00
	highest	City 2.06	Flat 0.57	High 0.96	Snow 1.20	Night 0.72	5.51
type a text message	lowest	Country 0.00	Trees 0.02	None 0.02	Rain 0.00	Day 0.03	0.07
	highest	City 1.85	Flat 0.76	Moderate 0.61	Snow 1.34	Night 0.10	4.66
watch a video	lowest	Country 0.00	Hills 0.00	None 0.08	Rain 0.00	Day 0.08	0.16
	highest	City 1.56	Trees 1.62	High 0.88	Snow 1.76	Night 0.93	6.75
Talk on the phone	lowest	Country 0.06	Hills 0.01	Moderate 0.01	Dry 0.00	Day 0.01	0.09
	highest	City 1.6	Flat 1.76	High 0.55	Rain 1.64	Night 0.70	6.25
make a shopping list	lowest	Country 0.06	Flat 0.00	None 0.00	Rain 0.22	Day 0.03	0.31
	highest	City 1.91	Hills 2.10	High 2.21	Dry 1.22	Night 2.19	9.63
adjust volume	lowest	Country 0.01	Hills 0.00	Moderate 0.03	Dry 0.00	Day 0.02	0.06
	highest	City 3.11	Trees 0.53	High 2.63	Snow 3.48	Night 2.32	12.07

4.4. Conclusion

The CBCA showed a secondary task modality effect for the factors and factor levels, indicating that there are relevant differences in the interaction of driving situation and secondary task.

Consistent with Fastenmeier [27], the context factor street type showed the highest impact on the demanded secondary task modality.

Lerner et al. [28] found participants to not attribute particular risk to basic use of smartphone functions, such as dialling, answering, and conversing. Here, *read a text message* and *type a text message* were found to be the least wanted to be executed in every driving situation, whereas *talk on the phone hands-free* was of high willingness.

As Huemer and Vollrath [39] observed, drivers more frequently engage in their smartphones when driving on a highway than when driving in a city. Going hand in hand with [52, 53] findings, drivers were less willing to use their smartphones on city roads. Carsten et al. [16] also found secondary task engagement most frequently on urban roads, and secondly on country motorway. Whereas little secondary task activity was found for rural roads, participants in the online survey were most willing to engage in secondary tasks on rural roads. Overall, country roads under dry weather were assigned the highest willingness to engage in a secondary task. In UDRIVE [16], country motorways in non-adverse weather conditions were identified as the most frequent context for secondary task engagement.

Supporting [52, 55] findings, willingness to engage in secondary tasks under bad weather, heavy traffic and at nights was low for all secondary tasks.

Female drivers were found to more frequently use their smartphones for texting and answering calls [51]. In contrast, [16] found women and men to be equally engaged in mobile phone tasks. No effect for gender was found here. Though, the willingness to engage in a secondary task was influenced by the age, annual mileage and driver profile. As previously found by [29, 40], age had a significant effect on the willingness to engage in some secondary tasks. Here, slightly younger participants were found to be more willing to *read a text message, talk on the phone hands-free*, *make a shopping list mentally* and *adjust the volume manually* in every driving situation. Further, annual mileage was found to influence the decision as well. *Watch a video* and *talk on the phone hands-free* was more likely for participants with a higher annual mileage. Only for *read a text message* an effect of driver profiles was found. Rather angry and dissociative classified drivers were more willing to read a text message in all of the driving situations than less careful and distress-reducing drivers.

5. Implications

Based on these findings, there is no single driving situation that has comparable effects on driver's perceptions whether to engage or not in a secondary task.

Although lock-outs were shown to have a positive effect on driving safety [41], participants of the present focus group reported to use alternatives to operate the restricted functions that are not safe for driving. As in [42], participants potentially experience a loss in autonomy, leading to psychological reactance.

Since the user requirements change over time [43, 44], influenced by the context [18, 19, 21], the user groups are heterogeneous [44] and changes in the environment [44, 45], adaptivity in the HMI is indicated (Fig. 17).

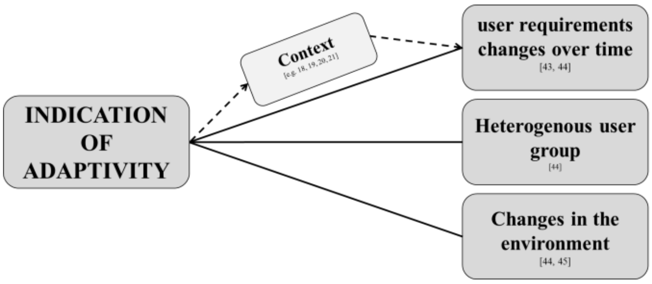

Fig. 17. *Indication of Adaptivity*

Depending on the driving situations, the HMI and warnings should be adapted accordingly. Whereas cognitive-auditory tasks seem to be unproblematic in most traffic scenarios, visual-manual tasks should be reduced in high workload scenarios, such as city drives in rain. Based on the Yerkes-Dodson-Law [46], adapting HMI content should not be reduced in all situations. In low arousing situations, such as country roads with no traffic, the HMI content can contain more information than in higher arousing situations, such as city drives with high traffic density.

Further, the adaptation should be as predictive as the driver's anticipation of the driving situation to provide not only user experience but safety for driving by a higher system understanding. The adaptation shall then follow the hysteresis principles [47].

6. Conclusion

It is known that any secondary task is distracting [4]. But it is also known, that drivers engage into them nonetheless [11, 13, 14, 15, 16], and therefore use compensatory strategies to reduce distraction [56]. Hence should a system not only support but nudge this compensatory behaviour.

As the creativity workshop and the focus group revealed, spending the driving time usefully was the main motivational factor for drivers to engage in their smartphones while driving. Both the focus group and the online survey confirmed [18, 19, 20, 21] findings on the context-depending changes of driver's needs and requirements.

Supporting results of Naturalistic Driving Studies [11, 16] and the literature [26, 52, 53, 54, 55] context factors were found to have different effects on the willingness to engage in the secondary task in question. The results of the online survey have the potential to quantify driving situations defined by the street type, landscape, traffic density, weather and daytime.

In order to gain insights on opinions and perception of the behaviour [17], the cascade of the explorative approach, consisting of creativity workshops, a focus group and an online survey, provided a feasible way to obtain a comprehensive understanding of driver needs and requirements in extending infotainment features. For automotive manufacturers, designing an infotainment system that fulfils both the need for information and reduction of distraction is desirable.

6.1 Limitations

The limited sample size and heterogeneity regarding company affiliations for the creativity workshops and the focus group raise caution regarding interpretation and generalisation of the findings.

Nonetheless, the results of the online survey are only subjective perceptions. Therefore, it is needed to further evaluate these findings in a simulator study, where a control on driving situation factor levels is possible. The here identified utilities of the CBCA can then be tested as predicting factors. Regarding the CBCA, some participants noted to be missing the alternative of *"I would never use the function."*. The secondary task m*ake a shopping list* showed to be unsuitable, since supermarkets in Germany close at latest at midnight, so making a shopping list at night apparently did not make sense to participants.

6.2 Further Research

Further investigations on drivers' behavioural adaptations in using their mobile devices when driving a car should be pursued.

In order to test the interaction of driving situation and secondary task, a driving study is needed to investigate the effects on the Collision Avoidance Metrics Programme [48], that is driving, glance and event detection behaviour, plus on subjectively perceived distraction and disturbance. As compensatory behaviour while engaged in secondary tasks in different driving situations is explored, the contradiction of wanting to be connected without being distracted, can be resolved by designing adapting current infotainment systems accordingly.

In addition, a real-driving study over a longer timeframe is recommendable to outline driving scenarios and investigate the adaptation based on these.

As [29, 49] found, legislation influences the perceived risk and willingness to engage in secondary tasks on mobile devices while driving. Therefore, further investigations on the legislation of mobile device usage while driving are needed.

Due to adaptations of current guidelines of driver distraction [50] regarding portable mobile devices, a continuous research investigating subjective user needs and requirements shall be pursued.

7. References

7.1. Bibliography

[1] mm1: 'Digitalisierung im Auto: Sitzt die deutsche Automobilindustrie in Zukunft noch am Steuer?' Online on the Internet, 2015, URL: https://mm1.de/fileadmin/content/Whitepaper/mm1_White_Paper_Digitalisierung_im_Auto_9_2015.pdf (Accessed 28 April 2017)

[2] Walker, G. H., Stanton, N. A., & Young, M. S.: 'Where is computing driving cars?'. International Journal of Human-Computer Interaction, 2001, 13(2), pp. 203-229.

[3] Vollrath, M., & Krems, J.: 'Fahren', in: Vollrath, M., & Krems, J. F. (Eds.): 'Verkehrspsychologie. Ein Lehrbuch für Psychologen, Ingenieure und Informatiker.' (Kohlhammer, 2012, 1st edn.), pp. 26-40

[4] Wickens, C. D.: 'Multiple resources and performance prediction. 'Theoretical issues in ergonomics science, 2002, 3(2), pp. 159-177

[5] Fuller, R.: 'The task-capability interface model of the driving process.' Recherche – Transports - Sécurité, 2000, (66), pp. 47-57

[6] Summala, H., Lamble, D., Laakso, M.: 'Driving experience and perception of the lead car's braking when looking at in-car targets.' Accident Analysis and Prevention, 1998, 30 (4), pp. 401–407.

[7] Reimer, B.: 'Impact of cognitive task complexity on drivers' visual tunnelling.' Trans. Res. Rec.: J. Trans. Res. Board, 2009, 2138, pp. 13–19

[8] Tsimhoni, O., Smith, D., Green, P.: 'Address entry while driving: speech recognition versus a touch-screen keyboard.' Hum. Factors J. Hum. Factors Ergonomics Soc., 2004, 46 (4), pp. 600–610.

[9] Engström, J., Johansson, E., & Östlund, J.: 'Effects of visual and cognitive load in real and simulated motorway driving.' Transportation research part F: Traffic psychology and behaviour, 2005, 8(2), pp. 97-120.

[10] AAA Foundation for Traffic Safety: '2014 Traffic Safety Culture Index.' Online on the internet, 2015, URL: https://www.aaafoundation.org/sites/default/files/2014TSCIreport.pdf (Accessed 18 May 2017)

[11] Dingus, T. A., Klauer, S. G., Neale, V. L., et al.: 'The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment (No. HS-810 593) '. National Highway Traffic Safety Administration, 2006

[12] World Health Organization: 'Mobile Phone Use: A Growing Problem of Driver Distraction.' WHO, 2011

[13] Dingus, T. A., Guo, F., Lee, S., et al.: 'Driver crash risk factors and prevalence evaluation using naturalistic driving data.' Proceedings of the National Academy of Sciences, 2016, 113(10), pp. 2636-2641

[14] Victor, T., Dozza, M., Bärgman, J., Boda, C. N., Engström, J., Flannagan, C., & Markkula, G.: 'Analysis of naturalistic driving study data: Safer glances, driver inattention, and crash risk.' SHRP 2 Report S2-S08A-RW-1, 2015

[15] Landau, A., Metz, B., Hargutt, V., et al.: 'Re-use of FOT data to assess the frequency and impact of distracted driving.' Presentation at the Naturalistic Driving Research Symposium, 2014, Blacksburg, Virginia, 26th of August 2014

[16] Carsten, O., Hibberd, D., Bärgman, J., Kovaceva, J., Pereira Cocron, M.S., Dotzauer, M. & Forcolin, F.: 'Driver Distraction and Inattention. UDRIVE Deliverable 43.1.'. EU FP7 Project UDRIVE Consortium, 2017

[17] Parnell, K. J., Stanton, N. A., & Plant, K. L.: 'What does current research tell us about why drivers engage with technological distractions: A Review.' 5th International Conference on Driver Distraction and Inattention, Paris, 2017

[18] Baltrunas, L., Kaminskas, M., Ludwig, B., et al.: 'InCarMusic: Context-aware music recommendations in a car.', in: Huemer, C. & Setzer, T. (Eds.) 'International Conference on Electronic Commerce and Web Technologies.', (Springer, 2011), pp. 89-100

[19] Dey, A. K.: 'Understanding and using context.' Personal and ubiquitous computing, 2001, 5(1), pp. 4-7

[20] Fischer, G.: 'Context-aware systems: the 'right' information, at the 'right' time, in the 'right' place, in the 'right' way, to the 'right' person.' Proceedings of the international working conference on advanced visual interfaces, 2012, pp. 287-294

[21] Helmholz, P.: 'Kontextorientierte Musikwiedergabe im Fahrzeug.', Doctoral dissertation, Technische Universität Braunschweig, 2015

[22] Parnell, K. J., Stanton, N. A., & Plant, K. L.: 'What technologies do people engage with while driving and why?.' Accident Analysis & Prevention, 2018, 111, pp. 222-237.

[23] Michalko, M.: 'Thinkertoys: A Handbook of Creative-Thinking Techniques' (Ten Speed Press, 2006, 2nd edn.)

[24] Krueger, R. A., & Casey, M. A.: 'Designing and conducting focus group interviews.' Social analysis, selected tools and techniques, 2002, 4(23), pp. 4-24

[25] Karrer, K., Glaser, C., Clemens, C., et al.: 'Technikaffinität erfassen–der Fragebogen TA-EG.' Der Mensch im Mittelpunkt technischer Systeme, 2009, (8), pp. 196-201

[26] Horberry, T., Anderson, J., Regan, M. A., Triggs, T. J. & Brown, J.: ‘Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance.’ Accident Analysis and Prevention, 2006, 38, pp. 185-191

[27] Fastenmeier, W.: ‘Autofahrer und Verkehrssituation. Neue Wege zur Bewertung von Sicherheit und Zuverlässigkeit moderner Strassenverkehrssysteme‘. Mensch, Fahrzeug, Umwelt, Bd. 33. (Verlag TÜV Rheinland, 1995)

[28] Lerner, N., Singer, J. P., & Huey, R.: ‘Driver strategies for engaging in distracting tasks using in-vehicle technologies (No. HS-DOT-810-919).’ United States. National Highway Traffic Safety Administration, 2008

[29] Parnell, K. J., Stanton, N. A., & Plant, K. L.: ‘Understanding drivers’ strategies for engaging with in-vehicle technology while driving: An interview study’ 5th International Conference on Driver Distraction and Inattention, Paris, 2017

[30] Taubman-Ben-Ari, O., Mikulincer, M., & Gillath, O.: 'The multidimensional driving style inventory - scale construct and validation.' Accident Analysis & Prevention, 2004, 36(3), pp. 323-332

[31] van Huysduynen, H. H., Terken, J., Martens, J. B., et al.: ‘Measuring driving styles: a validation of the multidimensional driving style inventory.’ Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, 2015, pp. 257-264

[32] Addelman, S., & Kempthorne, O. (1961). ‘Orthogonal main-effects plans.’ Ames, Iowa State Universtity: Aeronautical Reserach Laboratory Office if Aerospace Research United States Air Force, 1996

[33] Rao, V. R.: ‘Applied conjoint analysis.’ (Springer, 2014)

[34] Stinson, D. R.: ‘Combinatorial Designs.’ (Springer, 2004)

[35] Backhaus, K., Erichson, B., & Weiber, R.: ‘Fortgeschrittene multivariate Analysemethoden: eine anwendungsorientierte Einführung.‘ (Springer, 2015)

[36] Djokic, N., Salai, S., Kovac-Znidersic, R., Djokic, I., & Tomic, G.: ‘The use of conjoint and cluster analysis for preference-based market segmentation.’ Engineering Economics, 2013, 24(4), pp. 343-355.

[37] Spiessl, W., & Hussmann, H.: ‘Assessing error recognition in automated driving’. IET Intelligent Transport Systems, 2010, 5(2), pp. 103-111

[38] DeSarbo, W. S., Ramaswamy, V., & Cohen, S. H.: ‘Market segmentation with choice-based conjoint analysis.’ Marketing Letters, 1995, 6(2), pp. 137-147.

[39] Huemer, A. K. & Vollrath, M.: ‘Literature review on observational studies on cell phone use while driving’ 5th International Conference on Driver Distraction and Inattention, Paris, 2017

[40] Lerner, N., & Boyd, S.: ‘Task Report: On-Road Study of Willingness to Engage in Distraction Tasks (No. DOT HS 809 863)’. National Highway Traffic Safety Administration, 2005

[41] Lee, J. D.: ‘Technology and teen drivers.’ Journal of safety research, 2007, 38(2), pp. 203-213.

[42] Jung, T., Kaß, C., Schramm, T., & Zapf, D.: ‘So what really is user experience? An experimental study of user needs and emotional responses as underlying constructs.’ Ergonomics, 2017, 60(12), pp. 1601-1620.

[43] Brusilovsky, P.; & Millán, E.: ‘User Models for Adaptive Hypermedia and Adaptive Educational Systems.’ In: Brusilovsky, P., Kobsa, A. & Nejdl, W. (Eds.): The Adaptive Web, (Springer, 2017), pp. 3 – 53.

[44] Duarte, C. A.: ‘Design and Evaluation of Adaptive Multimodal Systems’, Doctoral Dissertation, Universidade de Lisboa, 2007

[45] Hastie, H. W., Johnston, M., & Ehlen, P.: ‘Context-sensitive help for multimodal dialogue.’ Proceedings of the Fourth IEEE International Conference on Multimodal Interfaces, 2002, pp. 93-98.

[46] Yerkes, R. M., & Dodson, J. D.: ‘The relation of strength of stimulus to rapidity of habit-formation.’ Journal of comparative neurology, 1908, 18(5), pp. 459-482.

[47] Farrell, P. S.: ‘The hysteresis effect.’ Human factors, 1999, 41(2), pp. 226-240.

[48] National Highway Traffic Safety Administration: ‘Visual-manual NHTSA Driver Distraction Guidelines for In-vehicle Devices. Docket No. NHTSA-2010-0053’, Washington, DC: National Highway Traffic Safety Administration (NHTSA), 2012

[49] Young, K. L., & Lenné, M. G.: ‘Driver engagement in distracting activities and the strategies used to minimise risk.’ Safety Science, 2010, 48(3), pp. 326-332.

[50] National Highway Traffic Safety Administration: ‘Visual-Manual NHTSA Driver Distraction Guidelines for Portable and Aftermarket Electronic Devices. Docket No. NHTSA–2013–0137’, Washington, DC: National Highway Traffic Safety Administration (NHTSA), 2014

[51] Oviedo-Trespalacios, O., Haque, M. M., King, M., & Washington, S.: ‘Should I Text or Call Here? A Situation-Based Analysis of Drivers’ Perceived Likelihood of Engaging in Mobile Phone Multitasking.’ Risk Analysis, 2018, DOI: 10.1111/risa.13119

[52] Britschgi V, Penttinen M, Rämä P.: 'Finnish Drivers and the use of In-Vehicle Technologies- Comparison of the focus group study and the internet survey.' In: Pace J, Mendoza L. (Eds.): 'Proceedings of European Conference on Human Centred Design for Intelligent Transport Systems', (HUMANIST Publications, 2012)

[53] Ferreira A, Rôla S, Simões A, Piccinini G.: 'Perceptions of Portuguese Drivers about the Usage of Mobile Phone While Driving.' In: Pace J, Mendoza L. (Eds.): 'Proceedings of European Conference on Human Centred Design for Intelligent Transport Systems', (HUMANIST Publications, 2012)

[54] Hancox, G., Richardson, J., & Morris, A.: 'Drivers' willingness to engage with their mobile phone: the influence of phone function and road demand.' IET Intelligent Transport Systems, 2013, 7(2), pp. 215-222

[55] Young, K. L., & Lenné, M. G.: 'Driver engagement in distracting activities and the strategies used to minimise risk.' Safety Science, 2010, 48(3), pp. 326-332

[56] Young, R. A.: 'Cell Phone Conversation and Relative Crash Risk Update.' in: Khosrow-Pour, M. (Ed.): Encyclopedia of Information Science and Technology.' (IGI Global, 2018, 4th edn.), pp. 5992-6006

7.2. Pictures Fig. 4

[1] Stokpic, https://www.pexels.com/de/foto/stadt-strasse-strasse-italien-5166/, accessed 15 August 2017

[2] Chensiyuan, https://upload.wikimedia.org/wikipedia/commons/e/e7/Quebec_City_Rue_St-Louis_2010b.jpg, accessed 15 August 2017

[3] Nout Gons, https://www.pexels.com/photo/action-america-architecture-avenue-378570/, accessed 15 August 2017

[4] Negative Space, https://www.pexels.com/photo/road-clouds-mountain-car-34585/, accessed 15 August 2017

[5] Josh Sorenson, https://www.pexels.com/photo/landscape-hilly-hills-highway-59512/, accessed 15 August 2017

[6] Canon PowerShot Sx200 Is, http://maxpixel.freegreatpicture.com/Winter-Landscape-Brandenburg-Tree-Lined-Avenue-Road-1156383, accessed 15 August 2017

[7] **S**andor Bordas, https://de.wikipedia.org/wiki/Datei:04.12.2007._Pyhrn-Autobahn_-_A_-_panoramio.jpg, accessed 15 August 2017

[8] Klaus Foehl, https://upload.wikimedia.org/wikipedia/commons/8/80/Autobahn_Tauberbischofsheim01_2009-04-13.jpg, accessed 15 August 2017

[9] RolandSD, https://pixabay.com/de/road-polen-spur-autobahn-umzug-1667218/, accessed 15 August 2017

8. Appendices

Table 8. Fractional 3^4 Design [24], Plan 3 [19] for driving situations

PROFILE	FACTOR A street	 B landscape	 C traffic density	 D weather	 E day time
1	A1 city	B1 flat	C1 none	D1 dry	Day
2	A1 city	B2 hills	C2 moderate	D3 snow	Day
3	A1 city	B3 trees	C3 high	D2 rain	Day
4	A2 country	B1 flat	C2 moderate	D2 rain	Day
5	A2 country	B2 hills	C3 high	D1 dry	Day
6	A2 country	B3 trees	C1 none	D3 snow	Day
7	A3 highway	B1 flat	C3 high	D3 snow	Day
8	A3 highway	B2 hills	C1 none	D2 rain	Day
9	A3 highway	B3 trees	C2 moderate	D1 dry	day
10	A1 city	B1 flat	C1 none	D1 dry	night
11	A1 city	B2 hills	C2 moderate	D3 snow	night
12	A1 city	B3 trees	C3 high	D2 rain	night
13	A2 country	B1 flat	C2 moderate	D2 rain	night
14	A2 country	B2 hills	C3 high	D1 dry	night
15	A2 country	B3 trees	C1 none	D3 snow	night
16	A3 highway	B1 flat	C3 high	D3 snow	night
17	A3 highway	B2 hills	C1 none	D2 rain	night
18	A3 highway	B3 trees	C2 moderate	D1 dry	night

Table 9. Path-worth utilities for each secondary task

		Read a text message	Type a text message	Watch a video	Phone call hands-free	Make a shopping list	Adjust music volume
Street type	City	2.06	1.85	1.56	1.61	1.91	3.11
	Country	0.00	0.00	0.00	0.07	0.06	0.01
	Highway	0.89	1.29	0.75	0.27	0.74	1.62
Landscape	Flat	0.57	0.76	0.99	1.76	0.00	0.13
	Hills	0.00	0.49	0.00	0.01	2.10	0.00
	Trees	0.26	0.02	1.62	0.08	0.93	0.53
Traffic density	None	0.11	0.02	0.08	0.21	0.00	1.87
	Moderate	0.00	0.61	0.38	0.02	0.51	0.03
	High	0.96	0.29	0.88	0.55	2.21	2.63
Weather	Dry	0.00	0.14	0.28	0.00	1.23	0.00
	Rain	0.71	0.00	0.00	1.64	0.23	1.37
	Snow	1.20	1.34	1.76	1.33	0.80	3.48
Day time	Day	0.00	0.04	0.08	0.01	0.04	0.02
	Night	0.72	0.10	0.93	0.70	2.19	2.32

Pre-crash driving behaviour of individuals with and without ADHD

Daniela Barragan [1*], Yi-Ching Lee [1]

[1] Department of Psychology, George Mason University, 4400 University Drive, Fairfax, USA
[*]dbarraga@gmu.edu

Abstract: Research suggests that drivers diagnosed with Attention-deficit/hyperactivity disorder (ADHD) are at increased risk of involvement in motor vehicle crashes due to inattention and impulsive behaviours. However, the behavioural characteristics of ADHD drivers which lead to a crash is not well understood. Therefore, the goal for this study was to evaluate the driving performance of individuals diagnosed with ADHD when they took their prescribed stimulant medication compared to when they refrained from taking their medication and a control condition. Forty-four participants (27 diagnosed with ADHD, 17 not diagnosed with ADHD) completed four simulated drives. ADHD drivers, when medicated, had similar pre-crash driving performance (velocity, brake force, steering movement, and lane offset) as the control condition. Conversely, when not medicated, ADHD drivers had significantly different driving performance compared to the medication and control conditions. These results highlight the importance that ADHD drivers take their medication, and noncompliance could be detected via in-vehicle safety systems.

1. Introduction

Motor vehicle crashes are the leading cause of death among young adults [1]. Young adult drivers diagnosed with attention-deficit/hyperactivity disorder (ADHD) are more likely to be involved in motor vehicle crashes than non-ADHD drivers [2, 3]. More specifically, Curry et al. (2017) found that young adult drivers diagnosed with ADHD are 36% more likely to be involved in a motor vehicle crash compared to drivers without ADHD.

Characteristics of ADHD individuals include inattention, impulsive behaviours, and unfocused motor activities [4]. One study found that driving performance of ADHD individuals was compatible to driving performance of intoxicated non-ADHD drivers [5]. Poor performance exhibited by ADHD drivers is due, in part, to deficits in cognitive functioning such as, difficulties attending to more than one object, poor speed management, inattention, and impulsivity [5, 6, 7]. Moreover, a recent study found that approximately 22% of motor vehicle crashes committed by ADHD drivers could have been prevented if they were medicated [2], suggesting that the cognitive deficits, which negatively impact ADHD drivers' performance may be mitigated through medication.

Stimulant medication prescribed to individuals diagnosed with ADHD has shown to increase arousal in these individuals [8]. While high levels of arousal are known to be detrimental to performance [9], a suitable increase in arousal from medication for ADHD individuals may likely lead to reduced driving impairment [10]. For example, Vaa (2014) suggests that ADHD drivers exhibit more speeding behaviour compared to non-ADHD drivers in an attempt to increase arousal [11]. Thus, medication may provide these individuals with an optimal level of arousal, which may consequently reduce the likelihood or severity of such unsafe driving behaviours.

Given the high prevalence of ADHD (4.40% of younger adults in the US) [12] and of preventable crashes among this population, it is important to further understand ADHD drivers' performance in relation to motor vehicle crashes. Specifically, the aim of this study was to evaluate performance differences between ADHD (when medicated and not medicated) and non-ADHD drivers prior to a crash to reveal which unsafe behaviours led to a crash. These results may also shed light on whether ADHD drivers are inherently unsafe drivers or if such detrimental behaviours can be remediated by medication. Therefore, the study's hypothesis was that medicated ADHD individuals would have similar driving performance prior to a crash as individuals without ADHD.

2. Method

2.1. Participants

Forty-four young drivers (17 without ADHD, 27 with ADHD) participated in the study. Participants were recruited from George Mason University and local communities through flyers and emails. All participants were between the ages of 18 and 24 (M = 20.82, SD = 1.79), held a valid US driving license, had normal or corrected-to-normal vision and hearing, and were either clinically diagnosed with ADHD or not. For the individuals with ADHD, their clinical diagnosis was verified via scores on Conners' Adult ADHD Rating Scales (CAARS) [13] and an ADHD symptoms survey. These participants were also required to be prescribed a Federal Drug Administration (FDA) approved simulant ADHD medication, which they took regularly (see [14]). Of the ADHD participants, 1 took Ritalin, 1 took Concerta, 4 took Adderall, 1 took Adderall XR, 2 took Vyvanse, 1 took Focalin, and 1 took Focalin XR (two participants took two medications). The non-ADHD participants were not clinically diagnosed with ADHD (verified via CAARS scores) nor did these individuals take ADHD medication.

Table 1 Means and standard deviations of survey scores for ADHD and control (non-ADHD) participants (n = 17)

Condition	Driving Anger Scale	Driving Behaviour Survey: Anxiety-based performance deficits	Driving Behaviour Survey: Exaggerated safety/caution behaviour	Driving Behaviour Survey: Hostile/aggressive behaviour	Brief Sensation Seeking Scale
Control	40.00 (7.46)	2.34 (.50)	4.97 (.91)	2.80 (1.10)	5.13 (4.12)
ADHD	43.67 (10.07)	3.00 (1.00)	4.76 (.44)	3.11 (.89)	10.44 (4.16)

Further details about participant screening and eligibility criteria were documented in the study protocol (see [15]).

Twenty-eight participants met the eligibility requirements. However, given that the goal of the study was to evaluate drivers' behaviour prior to a crash, only participants who were involved in an at-fault crash during the experiment were included. Data from 17 participants (5 men, 3 women without ADHD; 6 men, 3 women with ADHD) were included in the present study. Participants were compensated at a rate of £21.29 ($30) per hour.

2.2. Materials

The experiment took place at George Mason University in a half-cab Realtime Technologies, Inc. motion-based high-fidelity driving simulator. The driving simulator was equipped with three cameras, which recorded participants foot movement, face, upper body, and over the shoulder view. The driving scenarios were programmed using Javascript, the driving environment was developed in SimVista and run using SimCreator. Participants completed a practice drive and four different experimental drives, each lasting between 7-15 minutes. The drives contained ambient traffic and consisted of one or two-lane roads in rural and urban environments. Additionally, throughout the experimental drives, participants encountered 50 unique events (drive one: 15 events, drive two: 10 events, drive three: 14 events, drive four: 11 events), which were previously developed and validated [15, 16, 17]. For example, some of the included events involved pedestrians or bicyclists unexpectedly crossing the road, construction zones, and lead vehicles braking abruptly. Most of the events required participants to perform a manoeuvre (e.g., braking, lane merge) in order to avoid a collision.

After completing the experimental drives, participants completed a series of surveys online via Qualtrics including demographics and driving history, Safe Speed Knowledge Test [18], Driving Behaviour Survey [19], Driving Anger Scale [20], and Brief Sensation Seeking Scale [21]. Research has suggested that these surveys and personality traits measured can discriminate between individuals with and without ADHD [22, 23]. For example, Lopez et al. (2015) found that ADHD individuals have higher scores of sensation seeking, which they suggest is a facet of impulsivity.

All participants also completed the CAARS [13] online via Multi Health Systems Assessments (MHS Inc.) and responded orally to the Simulator Sickness Screening [24]. Individuals with ADHD also completed the Conners' Adult ADHD Diagnostic Interview for DSM-IV (CAADID) [25] orally, and an ADHD symptoms survey via Qualtrics. Individuals with ADHD had a family member or a friend complete the observer-version of the CAARS. Scores on the CAARS (self and observer) and ADHD symptoms survey were used as an index of ADHD. In support, previous research has identified the CAARS as being a reliable and valid index of ADHD [13].

2.3. Procedure

The study procedures were approved by George Mason University Institutional Review Board (IRB) and all participants signed an informed consent form. Participants first completed the Simulator Sickness Screening, then completed the simulator drives, and finally completed the remaining self-report measures. The ADHD participants were medicated prior to completing the self-report measures. For the simulator drives, participants were instructed to drive as they normally would, remain in the right lane, and follow traffic and speed limit signs, and navigation instructions.

A number of safety measures were in place: ADHD participants were dropped off and picked up by a friend or family member, their medication intake was monitored, and participant safety was actively monitored during simulator driving by a researcher. ADHD participants were required to bring their stimulant ADHD medication in the correct prescription bottle. The researcher confirmed that the name on the prescription bottle matched that of the participant. All ADHD participants completed the study across two days; one day for the medicated condition and the other for the non-medicated condition. The order of medication conditions (ADHD participants) and drives were randomly counterbalanced across participants. In the non-medicated condition, participants did not take their ADHD medication the day of participation whereas, in the medicated condition, participants took their ADHD medication under experimenter supervision and waited one hour for the medication to take effect prior to completing the study.

Participants without ADHD completed the same simulator drives as ADHD participants, but they completed a shorter list of self-report measures (i.e., did not complete ADHD symptoms survey or CAADID). The study lasted two hours for participants without ADHD and five hours (across two days) for ADHD participants.

Finally, ADHD participants were asked to identify someone close to them (hereon referred to as observers) to complete two surveys (CAARS and ADHD symptoms survey) about the participant. These observers completed the surveys online or over the phone. An independent licensed clinical psychologist confirmed ADHD diagnoses by evaluating participant and observer responses on the CAARS and ADHD symptoms survey. ADHD participants with self-report or observer-report t-scores less than 60 on the CAARS were classified as not having ADHD and were ineligible to participate. Additionally, non-ADHD participants who had a self-report t-score greater than or equal to 60 on the CAARS were ineligible to participant.

3. Results

Table 2 Means and standard errors of pre-crash and crash data across conditions (control, medicated, non-medicated)

Condition	Sample	Velocity (m/s)	Steering angle (degrees)	Brake force (Newtons)	Lane offset (m)
Control	Pre-crash	15.24 (.09)	62.42 (.66)	18.21 (.78)	.32 (.004)
	Crash	7.36 (.30)	55.34 (1.19)	94.16 (4.05)	.25 (.018)
ADHD medicated	Pre-crash	15.17 (.09)	54.50 (.24)	23.18 (.89)	.33 (.004)
	Crash	4.78 (.20)	54.17 (.53)	106.08 (2.95)	.37 (.006)
ADHD non-medicated	Pre-crash	14.28 (.10)	48.32 (.50)	26.77 (.78)	.50 (.006)
	Crash	8.93 (.17)	44.11 (1.39)	62.91 (2.99)	.41 (.014)

The three experimental conditions were: control condition (non-ADHD participants), medicated condition, and non-medicated condition. Table 1 provides the means and standard deviations for ADHD and non-ADHD participants' scores on the Driving Anger Scale, the three Driving Behaviour Survey subscales (i.e., anxiety-based performance deficits, exaggerated safety/caution behaviour, and hostile/aggressive behaviour), and the Brief Sensation Seeking Scale. Using R, the results of an independent-samples t-test revealed that ADHD participants had significantly greater scores on the Brief Sensation Seeking Scale compared to the non-ADHD participants, $t(15) = 2.64$, $p = .02$. However, there were no significant differences in survey scores between ADHD and non-ADHD participants on the Driving Behaviour Survey and the Driving Anger Scale, *p*s > .05.

Driving data were recorded at 60 Hz. Among the variables recorded, this study evaluated velocity (m/s), brake force (Newtons), steering angle (absolute value in degrees), and lane offset (absolute value in metres from lane centre). MATLAB was used for data reduction and all statistical analyses were performed using R. Data were evaluated in terms of pre-crash and crash data. Pre-crash data were defined as five seconds prior to each crash sample. A crash was defined as occurring when the participant vehicle was less than or equal to two metres from another vehicle. Table 2 lists the means and standard errors of pre-crash and crash data across conditions.

On average, individuals with ADHD were involved in 3.44 (SD = 2.88, range: 1-9) crashes and individuals without ADHD were involved in 1.75 (SD = 1.39, range: 1-5) crashes. A two-samples Welch t-test showed that there were no significant differences between the mean number of crash for ADHD and non-ADHD participants, $t(11.82) = 1.57$, $p = .14$. Non-medicated ADHD (M = 3.00, SD = 2.10) participants were involved in more crashes than medicated ADHD (M = 1.63, SD = .74) and control participants.

Linear mixed effects models with a random intercept of subject type (ADHD, non-ADHD) nested within subject were performed using the lme4 package in R [26] to evaluate the effects of experimental condition (non-medicated, medicated, control) on velocity, brake force, steering, and lane offset prior to a crash. Standard errors for the mixed effects models were calculated using the sjstats package in R [27]. Post-hoc analyses with pairwise adjustments were also performed in R using the lsmeans package [28].

There was a significant effect of condition on velocity, $\beta = 1.91$, $SE = .017$, $p < .001$. Specifically, velocity was significantly lower prior to a crash in the non-medicated condition compared to the medicated condition, $\beta = -1.91$, $SE = .13$, $p < .001$. There were no significant differences in velocity between the ADHD (medicated and non-medicated) and control conditions, *p*s > .05.

There was a significant effect of condition on brake force, $\beta = -7.47$, $SE = .018$, $p < .001$. Specifically, prior to a crash, the non-medicated condition had significantly greater brake force compared to the medicated condition, $\beta = 7.59$, $SE = 1.29$, $p < .001$. There were no significant differences in brake force between the ADHD (medicated and non-medicated) and control conditions, *p*s > .05.

Steering movement was significantly different between conditions, $\beta = 9.04$, $SE = .018$, $p < .001$. The non-medicated condition had significantly reduced steering movement prior to a crash compared to the medicated ($\beta = -9.08$, $SE = .80$, $p < .001$) and control ($\beta = -13.94$, $SE = 3.89$, $p = .003$) conditions. Steering did not significantly differ pre-crash between the medicated and control conditions, $p = .23$.

Finally, there was a significant effect of condition on lane offset, $\beta = -.16$, $SE = .016$, $p < .001$. The non-medicated condition had significantly greater lane offset than the medicated condition, $\beta = .16$, $SE = .007$, $p < .001$. Lane offset did not significantly differ between ADHD (non-medicated and medicated) and non-ADHD drivers, *p*s > .05.

4. Conclusion

The current research, contrary to some prior simulator studies [3, 29] revealed that ADHD drivers were just as likely as non-ADHD drivers to be involved in a simulated crash. Medicated ADHD drivers exhibited behaviours (velocity, steering, brake, lane offset) similar to those of non-ADHD drivers. Additionally, it was found that prior to a crash, non-medicated ADHD drivers had significantly lower velocity, increased brake force, decreased steering movement, and increased lane offset compared to medicated ADHD drivers. The non-medicated ADHD drivers also had significantly less steering movement prior to a crash compared to the non-ADHD drivers.

The results that non-medicated ADHD drivers had reduced velocity and increased brake force prior to a crash could suggest that they were aware of an increased likelihood of a crash. However, when the crash occurred, these participants only increased their brake force by 135% whereas, the non-ADHD (417.08% increase) and medicated ADHD (357.64% increase) participants applied the brake more forcefully during a crash. It is possible that the non-medicated ADHD participants incorrectly estimated the type of manoeuvre necessary to avoid a crash. For example, given that non-medicated participants had increased brake force prior to a crash, but had the lowest percent increase in brake force during a crash suggests that when not medicated ADHD drivers may underestimate the required stopping distance. Alternatively, the non-medicated ADHD drivers could have been aware of impaired driving performance when not medicated and thus attempted to drive more cautiously. For example, research has shown that ADHD drivers are more likely to speed than non-ADHD drivers [3] which is why these individuals had reduced velocity prior to a crash.

Further, in comparison to the non-ADHD participants and the medicated ADHD participants, the non-medicated ADHD drivers significantly reduced their steering movement prior to a simulator crash. Since the non-medicated ADHD drivers had reduced velocity, it is plausible that less steering movement was required.

Likewise, the non-medicated ADHD drivers had increased lane offset prior to a crash compared to the medicated ADHD drivers. Similarly, Kingery et al. (2015) found that ADHD drivers, when not medicated exhibited greater lane position variability compared to non-ADHD drivers [30]. The results of the current study lend further support that non-medicated ADHD drivers perform inadequate driving manoeuvres and may have prioritized velocity and lane offset rather than brake force and steering. In support, a recent review article found that 78.57% of the studies reviewed provide evidence of the benefits of ADHD drivers taking stimulant ADHD medication [31]. Specifically, these individuals, when medicated, had improved steering and braking to sudden events.

Cox et al. (2008) evaluated the extent that various stimulant ADHD prescription drugs affect driver impairment. Specifically, there were no differences in driving impairment for ADHD drivers who were prescribed Concerta compared to those prescribed Adderall XR [32]. Although possible, it is unlikely that the various stimulant medications prescribed to the ADHD participants in the current study had differing effects on driving performance.

The results of the current study could suggest that when not medicated, individuals with ADHD exhibit more impulsive behaviours than when properly medicated causing them to either underestimate the likelihood of a crash or overestimate their ability in preventing a crash [29, 33]. In support of the latter, Fuermaier et al. (2017) suggest that individuals with ADHD are subject to a positive illusory bias whereby these individuals tend to overestimate their driving ability due to difficulties in introspection [33]. Likewise, research suggests that ADHD drivers exhibit strong beliefs of self-efficacy [3, 29]. Oftentimes, such beliefs coupled with the inherent impulsive behaviours characterized by ADHD, leads these individuals to terminate medication and treatment [3].

Future research should evaluate these performance measures in terms of variability, which may be more sensitive to subtle changes to reflect impaired driving performance. Additionally, measuring standard deviation of lateral position may further reveal whether ADHD drivers tend to deviate from their mean position or the centre of the lane.

Understanding the driving behaviour of individuals with ADHD prior to a motor vehicle crash may assist in developing mitigation techniques to reduce unsafe driving. For example, assistive in-vehicle technologies could be used to determine when individuals have not taken their medication or when their medication has worn off by assessing real-time changes in driving behaviour.

5. Acknowledgments

The authors would like to express appreciation for the support of the National Science Foundation [Award Number = 1653624].

6. References

[1] Liu, Y., Singh, S., Subramanian, R.: ‘Motor vehicle traffic crashes as a leading cause of death in the United States, 2010 and 2011’, Traffic Safety Facts Research Note, 2015, (Report No. DOT-HS-812-203), Washington, DC: National Highway Traffic Safety Administration

[2] Chang, Z., Quinn, P. D., Hur, K., et al.: ‘Association between medication for attention-deficit/hyperactivity disorder and risk of motor vehicle crashes’, JAMA Psychiatry, 2017, 74, (6), pp. 597-603

[3] Groom, M. J., van Loon, E., Daley, D., et al.: 'Driving behaviour in adults with attention deficit/hyperactivity disorder’, BMC Psychiatry, 2015, 15, (175), pp. 1–11

[4] The National Institute of Mental Health: ‘Attention deficit hyperactivity disorder’, 2016, Retrieved 2018 from https://www.nimh.nih.gov/health/topics/attention-deficit-hyperactivity-disorder-adhd/index.html

[5] Weafer, J., Camarillo, D., Fillmore, M. T., et al.: ‘Simulated driving performance of adults with ADHD: comparisons with alcohol intoxication’, Experimental and Clinical Psychopharmacology, 2008, 3, (16), pp. 251-263

[6] Curry, A. E., Metzger, K. B., Pfeiffer, M. R., et al.: ‘Motor vehicle crash risk among adolescents and young adults with attention-deficit/hyperactivity disorder’, JAMA Pediatr., 2017, 171, (8), pp. 756-763

[7] Reimer, B., Mehler, B., D’Ambrosio, L. A., et al.: ‘The impact of distractions on young adult drivers with attention deficit hyperactivity disorder (ADHD)’, Accident Analysis & Prevention, 2010, 42, (3), pp. 842-851

[8] Maul, J., Advokat, C.: ‘Stimulant medications for attention-deficit/hyperactivity disorder (ADHD) improve memory of emotional stimuli in ADHD-diagnosed college students’, Pharmacol. Biochem. Behav., 2013, 105, pp. 58-62

[9] Pavlidis, I., Dcosta, M., Taamneh, S., et al.: ‘Dissecting driver behaviours under cognitive, emotional, sensorimotor, and mixed stressors’, Scientific Reports, 2016, 25651, (6), pp. 1-12

[10] Randell, N. J. S., Charlton, S. G., Starkey, N. J.: ‘Driving with ADHD: performance effects and environment demand in traffic’, Journal of Attention Disorders, 2016

[11] Vaa, T.: ‘ADHD and relative risk of accidents in road traffic: a meta-analysis’, Accident Analysis & Prevention, 2014, 62, pp. 415-425

[12] Kessler, R. C., Adler, L., Barkley, R., et al.: 'The prevalence and correlates of adult ADHD in the United States: results from the national comorbidity survey replication', Am. J. Psychiatry, 2006, 4, (163), pp. 716-723

[13] Conners, C. K., Erhardt, D., Epstein, J. N., et al.: 'Self-rating of ADHD symptoms in adults: I. factor structure and normative data', J. Attention Disorders, 1999, 3, (3), pp. 131-151

[14] 'FDA Drug Safety Communication: Safety Review Update of Medications used to treat Attention-Deficit/Hyperactivity Disorder (ADHD) in adults, https://www.fda.gov/Drugs/DrugSafety/ucm279858.htm#SA, accessed 3 May 2018

[15] Lee, Y.-C., Ward McIntosh, C., Winston, F. K., et al.: 'Design of experimental protocol to examine medication non-adherence among young drivers diagnosed with ADHD: a driving simulator study', Contemporary Clinical Trials Communications, 2018, 11, pp. 149-155

[16] Lee, Y.-C., Winston, F. K.: 'Stress induction techniques in a driving simulator and reactions from newly licensed drivers', Transportation Research Part F: Psychology and Behaviour, 2016, 42, pp. 44-55

[17] McDonald, C. C., Kandadai, V., Loeb, H. S., et al.: 'Simulated driving assessment (SDA) for teen drivers: results from a validation study', Injury Prevention, 2015, 21, pp. 145-152

[18] Legree, P. J., Heffner, T. S., Psotka, J., et al.: 'Traffic crash involvement: experiential driving knowledge and stressful contextual antecedents', Journal of Applied Psychology, 2003, 88, (1), pp. 15-26

[19] Clapp, J. D., Olsen, S. A., Beck, J. G., et al.: 'The driving behaviour survey: scale construction and validation', Journal of Anxiety Disorders, 2011, 25, pp. 96-105

[20] Deffenbacher, J., L., Thwaites, G. A., Wallace, T. L., et al.: 'Social skills and cognitive-relaxation approaches to general anger reduction', Journal of Counseling Psychology, 1994, 41, pp. 386-396

[21] Hoyle, R. H., Stephenson, M. T., Palmgreen, P., et al: 'Reliability and validity of a brief measure of sensation seeking', Personality and Individual Differences, 2002, 32, (3), pp. 401-414

[22] Lopez, R., Dauvilliers, Y., Jaussent, I., et al.: 'A multidimensional approach of impulsivity in adult attention deficit hyperactivity disorder', Psychiatry Res., 2015, 227, (2-3), pp. 290-295

[23] Pironti, V., Lai, M., Müller, U., et al.: 'Personality traits in adults with attention-deficit hyperactivity disorder and their unaffected first-degree relatives', BJPsych Open, 2016, 2, (4), pp. 280-285

[24] Hoffman, R. B., Molino, J. A., Inman, V. W.: 'Driving simulator sickness management at Turner-Fairbank Highway Research Center', Driving Simulation Conference, (ISSN 1546-5071)

[25] Epstein, J. N., Johnson, D., Conners, C. K.: 'Conners' adult ADHD diagnostic interview for DSM-IV', Multi-Health Systems Inc., 2000

[26] Bates, D., Maechler, M., Bolker, B., et al.: 'Fitting linear mixed-effects models using lme4', Journal of Statistical Software, 2015, 67, (1), pp. 1-48

[27] Lüdecke, D.: 'Statistical functions for regression models', 2018

[28] Lenth, R. V.: 'Least-squares means: the R package lsmeans', Journal of Statistical Software, 2016, 69, (1), pp. 1-33

[29] Knouse, L. E., Bagwell., C. L., Barkley, R. A., et al.: 'Accuracy of self-evaluation in adults with ADHD: evidence from a driving study', J. Attention Disorders, 2005, 4, (8), pp. 221-234

[30] Kingery, K. M., Narad, M., Garner, A. A., et al.: 'Extended visual glances away from the roadway are associated with ADHD- and texting-related driving performance deficits in adolescents', J. Abnorm. Child Psychol., 2015, 43, (6), pp. 1175-1186

[31] Surman, C. B. H., Fried, R., Rhodewalt, L., et al.: 'Do pharmaceuticals improve driving in individuals with ADHD? A review of the literature and evidence for clinical practice', CNS Drugs, 2017, 31, (10), pp. 857-866

[32] Cox, D. J., Moore, M., Burket, R., et al.: 'Rebound effects with long-acting amphetamine or methylphenidate stimulant medication preparations among adolescent male drivers with attention-deficit/hyperactivity disorder', Journal of Child and Adolescent Psychopharmacology, 2008, 18, (1), pp. 1-10

[33] Fuermaier, A. B. M., Tucha, L., Evans, B. L., et al.: 'Driving and attention deficit hyperactivity disorder', Journal of Neural Transm., 2017, 124, (1), pp. 55-67

Prevalence and self-regulation of drivers' secondary task engagement at intersections: An evaluation using naturalistic driving data

Rashed Ismaeel*, Daryl Hibberd, Oliver Carsten

Institute for Transport Studies, University of Leeds, Leeds, UK
*ml11rai@leeds.ac.uk

Abstract: On the basis of naturalistic driving data, this study examined the prevalence of secondary task engagement at intersections and investigated how drivers self-regulate and manage such activities in accordance with changing roadways and demand situations. Video recordings were viewed to identify secondary tasks in which drivers engaged and situational factors, specifically those related to the complexity of driving situations. Results showed that one-third of the total intersection time was allocated to secondary task engagement and that greater engagement occurred at upstream and downstream areas of intersections than at areas falling within intersections. Drivers tended to more frequently engage in secondary tasks when their vehicles were stationary than when the vehicles were moving. Elderly drivers were less likely to engage in secondary tasks than younger drivers. Finally, drivers were less likely to engage in secondary tasks when they did not have priority than when they had priority and at intersections managed through traffic signs than in those controlled by traffic lights. In conclusion, drivers appear to engage selectively in secondary tasks at intersections in accordance with changes in the demands imposed by driving and roadway situations. In such circumstances, drivers likely respond to increased demand and reduce secondary task engagement to preserve processing resources. The findings offer the preliminary information necessary to develop driver training/education and awareness programmes on managing distractions and safe driving strategies.

1. Introduction

Driving is regarded as a complex multitasking activity that necessitates the simultaneous execution of several physical, cognitive and sensory skills. Despite such complexity, however, drivers commonly occupy themselves concurrently with distracting activities (secondary tasks) whilst driving [1]. Driver distraction can be defined as the diversion of attention away from safety-critical driving activities towards a competing activity, which may lead to insufficient or no attention being given to activities critical for safe driving [2]. Driver distraction is widely classified as a significant contributor to road crashes and a major concern for traffic safety [3-8].

Driving behaviours in the real world are illuminated using observational studies called Naturalistic Driving Studies (NDS), wherein data are collected through unobtrusive equipment that is installed in a vehicle, with no experimental intervention applied [9]. Previous NDSs provided sophisticated insights into the mechanisms that underlie the driver distraction-related process. An example is the Strategic Highway Research Program 2 project, which showed that drivers who engage in secondary tasks are exposed to double the risk of crashing than that presented to attentive drivers [3]. Notwithstanding the value provided by such initiatives, evaluating the crash risk arising from the performance of secondary activities without considering how drivers manage or self-regulate secondary task engagement does not unravel the entirety of the complexity that characterises safe driving.

The management of secondary task engagement encompasses decisions on when to engage in secondary activities, what types of activities to engage in and whether adjustment is to be made in accordance with variations in the demands imposed by the primary driving task [10]. Management can take place when drivers forgo secondary task engagement altogether whilst driving or when they refrain from engaging in specific secondary tasks under certain demanding situations. Acquiring a better understanding of secondary task management can improve estimations of crash risk and advance comprehension of the safety effects of driver distraction [11].

Many studies have implemented the naturalistic driving (ND) approach to illustrate how drivers manage their engagement in secondary tasks. An early study conducted in the US, for instance, revealed that drivers tend to less frequently occupy themselves with secondary tasks when they are driving at night, braking, driving on wet roadways and travelling through horizontal curves [12]. Other studies found that drivers are more likely to engage in secondary tasks when their vehicles are stationary than when they are moving [13-15]. In a similar study carried out in the Swedish context, the researchers concluded that drivers are less likely to initiate visual–manual secondary tasks during high-driving-demand situations (e.g. sharp turns and high speeds) and that drivers wait until they have completed lane changing manoeuvres before initiating secondary tasks [16]. These findings suggest that drivers self-regulate their behaviours by selecting what they evaluate as safe periods at which to engage in secondary tasks. A deficiency in this regard is the lack of studies that deal with drivers' self-regulation at intersections.

Focusing on intersections as one of the most safety-critical and highly demanding locations within a road network is a relevant component of enquiries into self-regulation behaviours because intersection negotiation poses extra demands on a driver and features heavily in crash statistics. For instance, intersection-related crashes represent nearly 50% of the total number of injury crashes in Germany [17] and nearly 60% of that in the UK [18]. Despite the fact that intersections are prominently implicated in crashes, there is a limited understanding of real-world driving behaviours at these locations. In particular, relatively little is known about the willingness of drivers to engage in secondary tasks and the manner by which they manage such engagement in accordance with changing demand situations at these sites.

The core idea that underpins this study is the in-depth analysis of drivers' engagement in secondary activities whilst performing manoeuvres at intersections. The analysis was based on ND data from the large-scale European naturalistic driving project known as the 'eUropean naturalistic Driving and Riding for Infrastructure & Vehicle safety and Environment' (UDRIVE). The importance of the current study lies in the combination of two key critical challenges to road safety: distractions and intersection. Correspondingly, the investigation was aimed at probing into the types of secondary tasks (e.g. mobile phone use, smoking, eating) that drivers typically engage in as they pass through intersections and exploring the prevalence of such conduct. The study was also intended to ascertain whether engagement in secondary tasks at intersections is influenced by driver-related personal characteristics, such as age and gender, and some situational variables, specifically those related to complex aspects of driving situations, including intersection control measures (traffic lights or traffic signs and road markings), intersection priority and vehicle status (moving or stationary). Finally, the study was directed towards a distraction-related comparison of the intersection approach phase (upstream functional area), the during-intersection phase (intersection physical area) and the beyond-intersection phase (downstream functional area) to explore how drivers manage secondary task engagement at intersections in accordance with changing roadways and demand situations.

2. Methods

To look into whether drivers adjust their secondary task engagement whilst driving at intersections, ND data from the UDRIVE project were coded and analysed. The observational data were supplemented by some driver-related factors (e.g. age and gender) which were obtained through questionnaires administered to participants in the recruitment stage. This study was approved by the Research Ethics Committee of the Environment Faculty at the University of Leeds (Ethics reference: AREA 16-193).

2.1. Participants

The sample comprised 163 car drivers (78 females and 85 males) who had more than 20 trips recorded in the UDRIVE dataset. Their age ranged from 18 to 80 years [mean = 43.8, standard deviation (SD) = 13.1] (Table 1), and their geographical locations were distributed across five European countries (the UK, France, the Netherlands, Germany and Poland).

Table 1. *Descriptive statistics of age groups (in years)*

Age	N	Mean	Minimum	Maximum	SD
18-29	19	24.7	18	28	3.2
30-39	52	34.6	30	39	3.0
40-59	66	48.0	40	58	5.6
60-80	26	65.3	60	80	5.7
Total	163	43.8	18	80	13.1

2.2. UDRIVE data acquisition system

The participants' own vehicles were equipped with a data acquisition system (DAS), which remained in the vehicles for 18 months from mid-2015 to early 2017. The DAS is composed of (1) a combination of sensors that automatically provide continuous measurements (e.g. an accelerometer, a global positioning system and an internal controller area network intended to measure speed, brake pedalling, engine revolutions per minute, etc.); (2) a smart forward-facing camera that detects and measures frontward distances from other road users; and (3) multiple other cameras for broad video coverage of road environments and driver behaviours (eight cameras in total) [19].

The cameras provide images of a driver's forward and side views and in-cabin views, with minimal disturbance to the driver's line of sight. These cameras are (1–3) three front cameras (left, centre, right) that capture approximately 180° views of a vehicle's front situation; (4) a face camera that captures a driver's facial expressions and gaze directions; (5) a blind spot camera that captures possible road users on the right side of a vehicle; (6) a driver action camera positioned over the shoulders to record the actions of a driver's hands; (7) a cabin camera that records the presence and activity of passengers; and (8) a foot camera that captures the actions of a driver's feet [20]. The participants could deactivate the recording system of the cameras by pressing a button assigned for this purpose [21]. This was considered a very important requirement for satisfying ethics standards and enabled the drivers to terminate recording temporarily for any reason as they drove.

2.3. Data sampling

The UDRIVE project yielded data on nearly 140,000 trips, with nearly 46,000 hours of ND data and nearly 1.5 million intersection cases. A robust sampling process was established for the selection of a representative sample of the UDRIVE dataset. The criteria used for sampling the intersection cases were as follows:

- A driver should have made at least 20 trips, with a minimum distance of 1 km per trip.
- For each driver, 10 trips were sampled randomly without replacement.
- For each trip, one intersection case was sampled randomly across all the intersection cases within that trip. Each intersection case was selected from a unique trip (no more than one intersection case selected per trip).
- For the annotated intersection cases, certain conditions had to be satisfied. That is, all camera channels should have been properly oriented and sufficient for annotation. Continuous measurements by sensors should have been existing and perfectly synchronised with the camera channels.

The above-mentioned selection process produced a sample of 163 drivers with 1630 intersection cases (10 intersection cases per driver).

2.4. Data coding and analysis

A scheme developed specifically for this study was used to code the selected sample to appropriately define different categories related to secondary activities, drivers' personal characteristics (age and gender) and situational factors. The key variables are described as follows:

Secondary tasks: The main dependent variable was the proportion of total intersection time accounted for by secondary tasks. Drawing from key distraction studies [10, 15, 22], the present research identified 10 secondary task types for annotation: passenger conversations (i.e. any instance of conversation, whether as minimal as a single-word utterance, with a passenger in an observed segment), talking/singing with no passengers present, mobile phone-related tasks, interactions with in-vehicle control systems (e.g. adjusting climate control), smoking-associated activities, grooming-related tasks, eating/drinking-related tasks, reading/writing-related activities, electronic device-related tasks and navigation system-related tasks (coded when a map can be observed from video channels or when some kind of interaction with a map transpires). A secondary task activity was annotated in a separate channel wherein multiple tasks could be coded simultaneously (e.g. a driver talking on a mobile phone whilst manipulating an in-vehicle control system).

Drivers' age categories: In line with the UDRIVE risk factors, crash causation and everyday driving reports [23], driver age was classified into four ordinal age groups: 18 to 29 years, 30 to 39 years, 40 to 59 years and 60 to 80 years.

Intersection type: Intersections were coded as roundabouts or intersections.

Intersection control: Intersection control measures were coded as control via traffic lights or management through traffic signs and road markings.

Intersection priority: Intersections were coded on the basis of priority as intersections in which a subject vehicle (SV) has priority or intersections in which an SV has no priority.

Turning direction: Three categories of turns were studied, namely, right turns, left turns and going straight through an intersection. A noteworthy point here is that the UK is the only country within the UDRIVE project where people drive on the left side of a road. Accordingly, descriptions of turning directions in the UK were flipped to match the data on the other countries.

Intersection locality: Intersection approaches were coded as located in urban or rural areas.

Vehicle motion status: The motion status of vehicles was coded as 'stationary' or 'moving' on the basis of time-series speed data. A stationary condition was defined as a situation wherein vehicle speed drops to zero (full stop). Earlier studies [e.g. 15] expected drivers to realise that the driving task is less demanding when vehicles are stationary and accordingly adapt secondary task engagement.

Given that the study was aimed at comparing secondary task engagement in the intersection upstream, intersection physical and intersection downstream areas, an essential requirement was to employ a mechanism that delineates the boundaries of these phases. The intersection functional area can be defined as a distance-based zone that extends both upstream and downstream beyond the boundaries of the intersection physical area [24]. The major component considered in determining this distance-based zone is the stopping sight distance (SSD). The SSD, in turn, is primarily based on speed and can be derived by adding the distance travelled during perception–reaction time to the distance travelled whilst braking [25–27].

The current work adopted the physical length values of the intersection functional area that were published in two previous studies (one step below the desirable minimum values) [24, 27]. These distance-based zone values were varied in accordance with the speed limit at intersections, as shown in Table 2.

Table 2. *The physical length of the intersection functional area as a function of speed at intersections*

Speed (km/h)	Physical length (m)
30	25
40	35
50	50
60	70
70	90
80	115
90	140
100	160

A UDRIVE-developed visualisation tool called SALSA (Smart Application for Large Scale Analysis) was used as the viewing platform to facilitate the viewing and annotation of the data. The reliability of the coded data was tested via inter-rater checks, for which a second independent coder coded 10% of the intersection cases. Inter-rater reliability was nearly 95% for the categorical variables (e.g. intersection priority) and nearly 90% for the continuous variables (e.g. secondary task duration).

The Statistical Package for the Social Sciences version 24 was used for the data analysis. Several descriptive and inferential analyses were carried out to examine the types and prevalence of secondary task engagement in relation to the selected situational and personal driver variables. The primary metric selected to evaluate the prevalence of secondary task engagement was the proportion of total intersection time accounted for by secondary tasks. Other metrics were the proportion of upstream intersection time, during-intersection time, downstream intersection time, total moving time and total stationary time accounted for by secondary tasks. The aforementioned variables were non-normally distributed; hence, non-parametric statistical tests were performed, namely, the Mann–Whitney U test, the Kruskal–Wallis H test, the Wilcoxon signed-rank test and the Friedman test.

3. Findings and discussion

The analysis was directed towards 1630 intersection cases, amounting to a total of 678.8 min of observation time. The mean duration of an intersection segment was 25 s. In term of motion status, the total 678.8 minutes observation time divided into 536 minutes of moving time and 142.8 minutes of stationary time. In term of the intersection stages, the total 678.8 minutes observation time divided into 373.2 minutes for upstream-, 161.2 minutes for during- and 144.4 minutes for downstream-stage.

With respect to the situational factors, the 1630 intersection cases were assigned to categories according to each situational variable in Table 3.

Table 3. *Situational factors obtained from data coding*

Situational factor	% of total intersection cases (1630 cases)
Intersection type	
Intersections	74.0
Roundabouts	26.0
Intersection control	
Traffic lights	37.1
Traffic signs and road markings	62.9
Intersection priority	
SV has priority	50.5
SV has no priority	49.5
Intersection locality	
Urban	75.3
Rural	24.7
Turning direction	
Turning right	32.5
Turning left	30.1
Going straight	37.4

3.1. Overall results on secondary task engagement

The analysis revealed that 50.9% of the intersection cases and 30.6% of the total intersection time involved engagement in at least one kind of secondary task. In other words, nearly one-half of the intersection cases and one-third of the total intersection time involved engagement in secondary tasks. Amongst all the cases, 555 (34.1%) were characterised by driver engagement in a single secondary task, 221 (13.6%) featured driver engagement in two secondary tasks and 53 (3.2%) involved driver engagement in more than two secondary tasks.

The UDRIVE project indicated that 52% of the coded trips and 10.2% of the analysed total travel time involved at least one secondary task [10]. By contrast, the current work discovered a higher level of secondary task engagement. This difference is likely due to coverage—the UDRIVE analysis was performed across the full range of driving contexts, whereas the present analysis was restricted to intersections. Moreover, the current study included two types of secondary tasks that were not covered in the UDRIVE project: passenger conversations and activities related to embedded/integrated in-vehicle navigation systems.

3.2. Frequency and prevalence of secondary tasks engagement

Table 4 presents the frequency at which the drivers performed each secondary task type at intersections. Overall, the data revealed that the most frequently observed tasks were passenger conversations (n = 456) and talking/singing in the absence of passengers (n = 148), followed by mobile phone interactions (n = 132), navigation system interactions (n = 109) and in-vehicle control system-related tasks (n = 99). Reading and writing tasks accounted for the lowest frequency (n = 6). These findings are consistent with previous NDSs in which passenger conversations were the leading secondary tasks observed [3, 15].

Table 4. *Frequency of secondary tasks*

Secondary task	Frequency
Passenger conversations	456
Talking/Singing with no passengers present	148
Mobile phone-related tasks	132
Navigation system-related tasks	109
Interactions with in-vehicle control systems	99
Smoking-related tasks	74
Personal grooming-related tasks	74
Food- and drink-related tasks	29
Reading- and writing-related tasks	6
Other	30
Total	1157

In the UDRIVE project, mobile phone usage and talking/singing tasks were the most frequent, whereas reading/writing was the lowest-frequency task [10], consistent with the findings of the current study (accounting for the absence of passenger conversation tasks). The only difference between the two studies is the relative frequency of personal grooming and food/drink-related tasks; that is, these behaviours occurred less frequently in the present study. This disparity may be attributed to the specificity of the driving manoeuvres executed in the intersections cases and may therefore suggest a form of self-regulatory practice by drivers. This self-regulation is that drivers ban or reduce their engagement in certain secondary tasks at intersections.

As a second step in analysing type of task engagement, the total amount of time that the drivers allocated to each secondary task was compared with the total intersection time (678.8 min). Figure 1 indicates that passenger conversations were the most frequently performed tasks, as determined from the proportion of these tasks out of the total intersection time (13.2%), followed by mobile usage (6.6%), navigation system-related tasks (6.6%) and smoking-related tasks (3.7%). The finding on passenger conversations accounting for the highest proportion of secondary task engagement is consistent with an early NDS in which passenger conversation was the leading secondary task, as determined on the basis of the proportion of time allocated to this task out of the total driving time (15.3%) [15]. Another NDS consistent with the current research showed that passenger conversations and mobile phone-related tasks accounted for 14.6% and 6.4% of the total baseline duration, respectively [3]. The results of the present study are also consistent with the findings on the UDRIVE general driving context, albeit this interpretation does not consider passenger conversations and navigation system-related tasks [10].

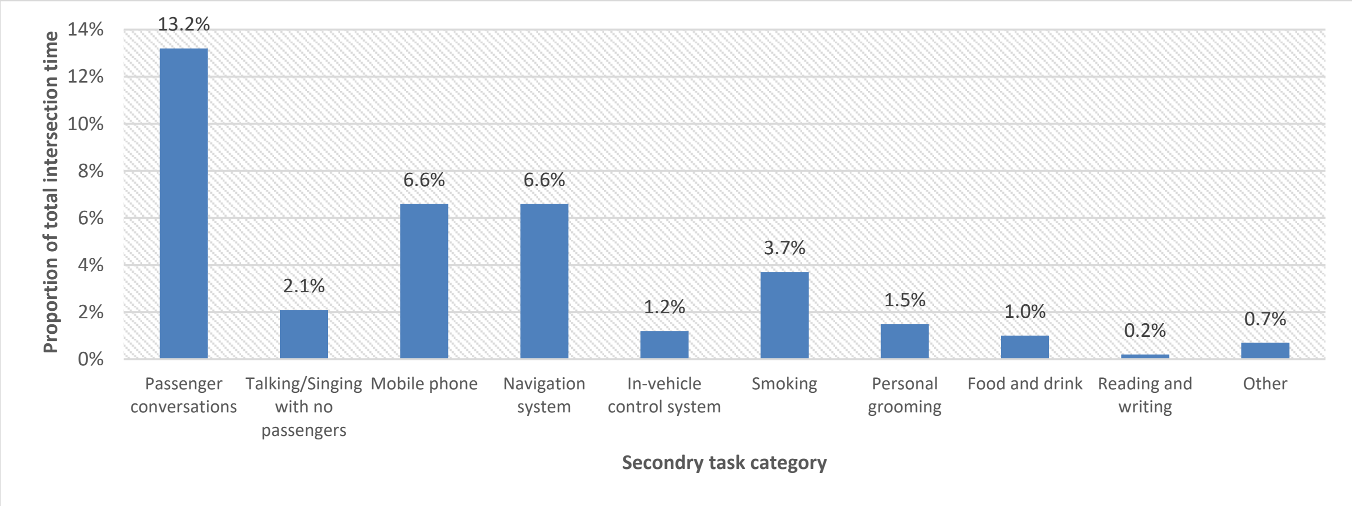

Figure 1. *Proportions out of total intersection time by type of secondary task*

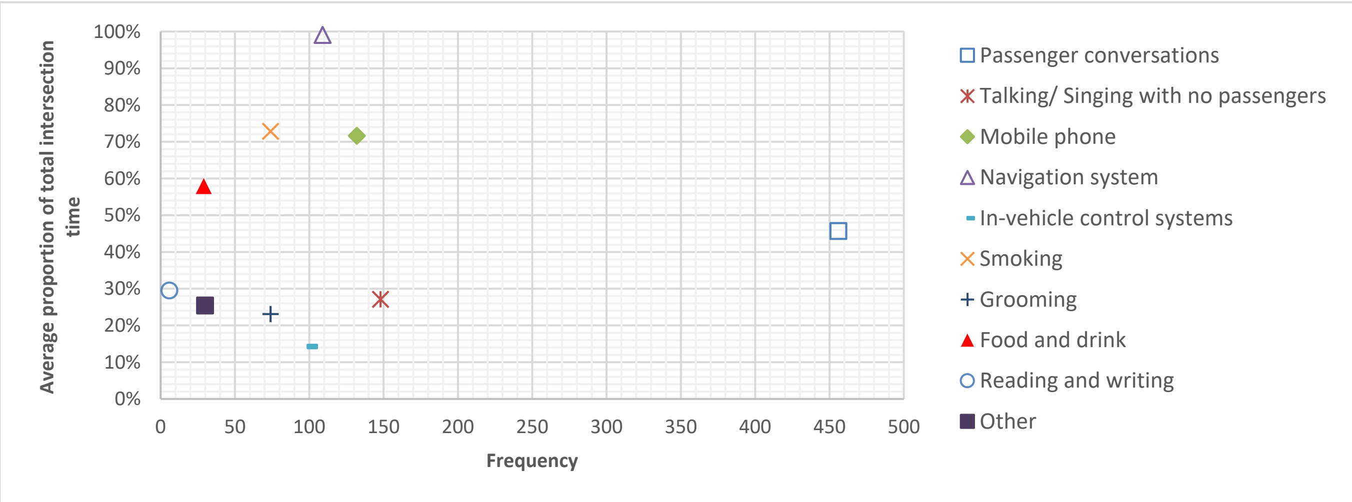

Figure 2. *Secondary task frequency versus average proportion of intersection time allocated to each task type*

The third step in delving into type of task engagement was determining the relationship between each task's frequency and average proportion accounted for in the total intersection time (Figure 2). Navigation system-related tasks were the activities to which the longest average proportion of time (99.1%) was devoted, which was an expected result given that such activities were coded for as long as a screen could be seen and independently from the modality of the interactions. Conversely, interactions with in-vehicle control systems were accorded the shortest average proportion of time (14.3%), which was also expected because of the short duration required to perform these tasks. Passenger conversations were devoted an average proportion of 45.7% but were by far the tasks with which the drivers most frequently occupied themselves. Mobile phone usage and smoking-associated tasks had similar average proportions, but the former was a more frequently exercised activity. Only six reading/writing tasks were observed within the annotated data, with these activities receiving an average proportion of 29.5% out of the total intersection time.

3.3. Proportions of time allocated to secondary tasks at each intersection stage and motion status

As mentioned earlier, 30.6% of the total intersection time was associated with secondary task engagement. Figure 3 shows a breakdown of the proportions of total intersection time by intersection stage (upstream, during and downstream) and by vehicle motion status (moving and stationary).

The Friedman test results showed that the proportions of time accounted for by secondary task engagement significantly differed at different intersection stages, $\chi2(2) = 76.364$, $p < 0.0005$. Pairwise comparisons indicated significantly higher engagement during the upstream and downstream stages than at the during intersection stage. These results suggest that drivers reduce secondary task engagement in the during-intersection stage as a self-regulatory measure. Drivers likely respond to increased risk/demand when they are located at the physical intersection area and reduce secondary task engagement to preserve processing resources. These outcomes constitute what can be called a V-shaped relationship between secondary task engagement and the three intersection stages (Figure 3).

The Wilcoxon signed-rank test revealed that the drivers significantly increased the proportion of time devoted to secondary tasks when their vehicles were stationary compared with when their vehicles were moving, $z = -7.196$, $p < 0.0005$. Again, the drivers appeared to self-regulate at intersections, indicating that they are more likely to perform secondary tasks when driving task demand is lower at stationary conditions (Figure 3). This increment in engagement, however, does not mean that it is a safe practice. In these situations, drivers will be compelled to generate extra cognitive load, which in turn, may slow down driver decision making.

3.4. Proportions of time allocated to secondary task engagement as determined by stationarity

Figure 4 illustrates the comparison of secondary task engagement at intersections with (436 cases) and without (1194 cases) stationarity. For both conditions, the Friedman test showed a significant difference in the proportions of time accounted for by secondary task engagement across the intersection stages. Moreover, the V-shaped relationship between secondary task engagement and intersection stage remained under the two scenarios (Figure 4).

Post-hoc pair-wise comparisons of the cases wherein no stationarity occurred revealed statistically significant increases in task proportions from the during-intersection to the downstream intersection stages ($p = 0.014$). In regard to the stationarity cases, the proportion of secondary task engagement was significantly higher in the upstream stage than in the during-intersection and downstream stages ($p < 0.0005$). These results suggest that drivers who pass through an intersection without stopping are more likely to wait until exit from the during-intersection stage before initiating secondary tasks. Those who stop, however, tend to allocate time to task engagement in the upstream stage and then abandon the activity to keep pace with the increasing demand encountered at the during-intersection stage. These results imply that stationarity significantly affects drivers' decisions regarding when to engage in secondary tasks across the intersection stages. Stationarity-induced engagement can thus be considered another form of self-regulatory practice.

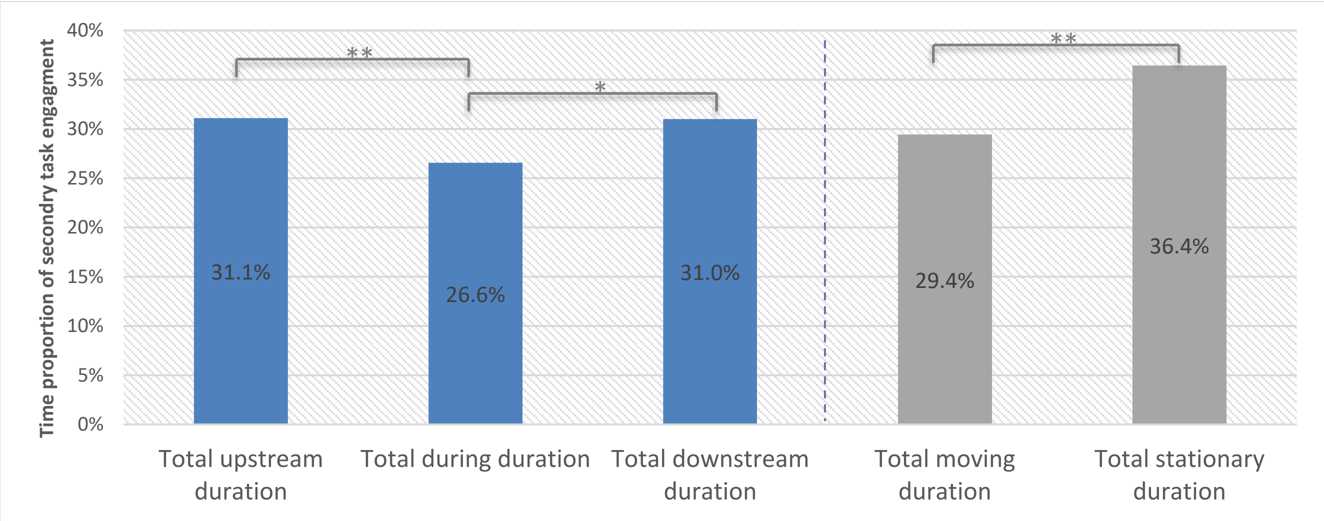

Figure 3. *Proportions of time allocated to secondary task engagement at each intersection stage and motion status (*p < 0.05, **p < 0.0005)*

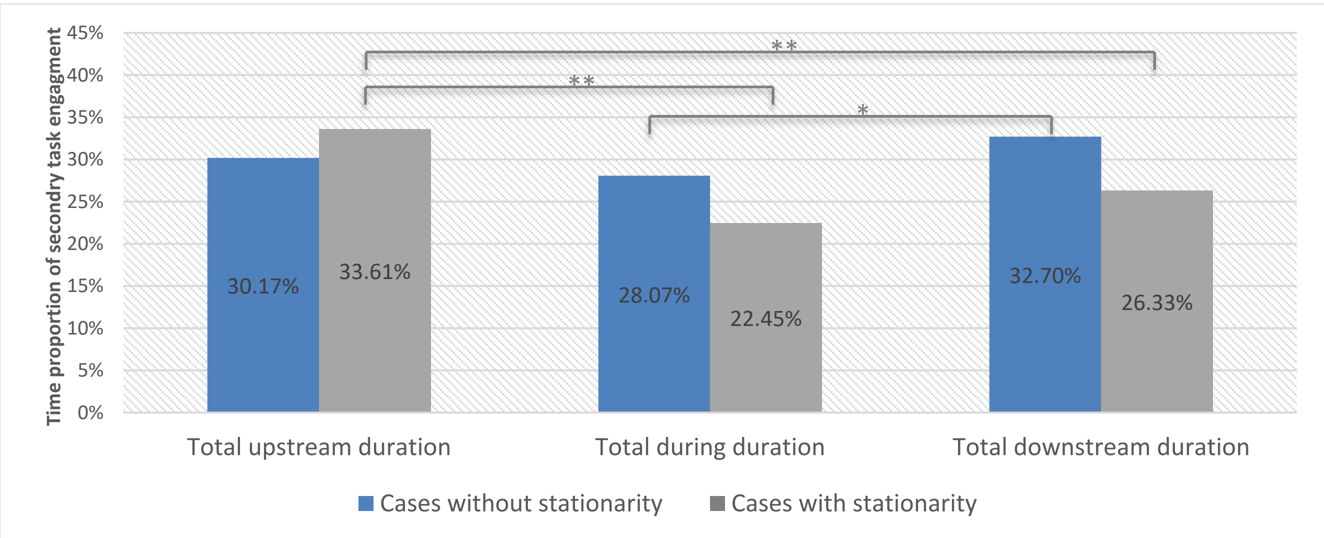

Figure 4. *Proportions of time allocated to secondary task engagement at each stage by presence of stationarity (*p < 0.05, **p < 0.0005)*

3.5. Secondary task engagement based on driver-related factors

The Mann–Whitney U and Kruskal–Wallis H tests were conducted to determine whether significant differences exist between the two genders and amongst the four age groups, respectively, in terms of the proportion of time allocated to secondary tasks. The time proportions tested here were those pertaining to the total intersection time, as well as the time proportions allocated to each intersection stage and motion status.

The data analysis revealed that gender did not exert a significant effect on the proportions of time that the drivers allocated to secondary task engagement (Table 5). This result is consistent with that of an Australian NDS, which reported no significant difference in task engagement allocations between males and females [14]. With respect to age, the analysis unravelled a significant difference in the proportions of time allocated to secondary task engagement amongst the age groups (Table 5). Proportion decreased with age under all the intersection stages and motion statuses. Figure 5 shows how the relationship between the proportion out of total intersection time and age group was shaped. The result indicates that elderly drivers are less likely to engage in secondary activities than younger ones—a finding that aligns with many studies within the driving literature [e.g. 15, 28]. A plausible conclusion, then, is that this behavioural trend is sustained, especially when one considers the complexity of driving at intersections and the reduced abilities of elderly drivers (e.g. sensory, cognitive and physical functioning).

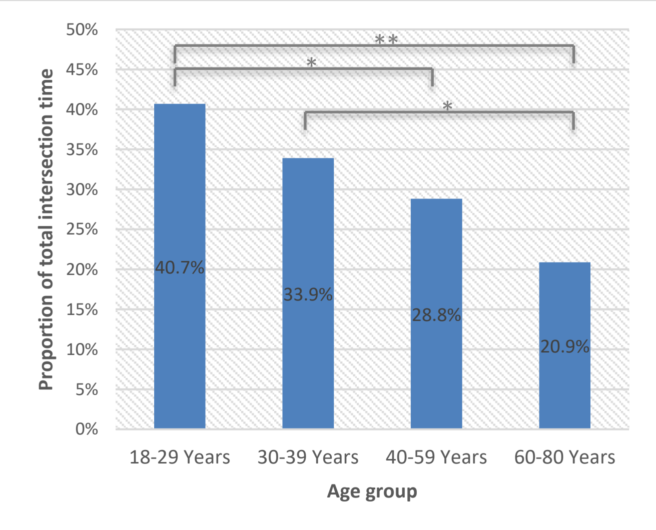

Figure 5. *Proportion of total intersection time allocated to secondary tasks by age group (*p < 0.05, **p < 0.0005)*

3.6. Secondary task engagement based on situational factors

The Mann–Whitney U test was used to determine whether significant differences exist amongst intersection type, intersection control, intersection priority and intersection locality with respect to the proportions of time allocated to secondary tasks. The Kruskal–Wallis H was carried out to determine whether significant differences exist amongst the three categories of turning directions in terms of the aforementioned proportions.

With regard to intersection control, the proportions of time out of the total intersection time ($z = –3.022$, $p = 0.003$), the total upstream time ($z = –4.498$, $p < 0.0005$) and the total stationary time ($z = –2.488$, $p = 0.013$) during which the drivers engaged in secondary tasks were lower at intersections with traffic signs than at intersections with traffic lights (Table 6). This result suggests that drivers self-regulate secondary task engagement by reducing interactions at intersections that are managed by traffic signs (which require gap judgments) to levels below engagement at intersections that are fully controlled by traffic lights.

With reference to intersection priority, the proportions of time in which the drivers were occupied with secondary tasks were significantly higher in situations wherein the drivers had priority, but this applied only with respect to total stationary time ($z = –3.005$, $p = 0.003$) (Table 6). This result suggests that drivers, whilst stationary condition at intersection approaches, are more likely to engage in secondary tasks when they have priority than when no such priority is accorded to them. This supposition is reasonable because drivers are required to judge gaps and choose the best option for crossing an intersection (high decision-making demand) as they stop at non-priority locations. Drivers therefore self-regulate secondary task engagement in these situations.

In terms of intersection type, the proportions of time out of the during time ($z = –2.110$, $p = 0.035$), the downstream time ($z = –2.241$, $p = 0.025$) and the total moving time ($z = –2.332$, $p = 0.020$) during which the drivers occupied themselves with secondary tasks were significantly higher at roundabouts than at intersections. This is a surprising result given that roundabouts are complex types of intersections that impose high driving task demands, especially at the during-intersection stage (circulating flow). Note, however, that possible confounding effects may have stemmed from the presence of stationarity, intersection priority and control factors in the comparison of roundabouts with other intersection types. Both intersection locality (urban/rural) and turning directions (left/right/straight) exerted no significant effect on the proportions of time that the drivers allocated to secondary task engagement (Table 6).

Table 5. *Presence of statistically significant differences in proportions of time allocated to secondary tasks based on driver-related factors (*p < 0.05, **p < 0.0005)*

Driver-related factor	% of total intersection time	% of upstream time	% of during time	% of downstream time	% of moving time	% of stationary time
Gender						
Age	**	**	**	**	**	*

Table 6. *Presence of statistically significant differences in proportions of time allocated to secondary tasks based on situational factors (*p < 0.05, **p < 0.0005)*

Situational factor	% of total intersection time	% of upstream time	% of during time	% of downstream time	% of moving time	% of stationary time
Intersection control	*	**				*
Intersection priority						*
Intersection type			*	*	*	
Intersection locality						
Turning direction						

4. Conclusion

This paper presents a novel application of the ND approach in the examination of driver engagement in secondary tasks at intersections. The findings on prevalence showed that 30.6% of the total intersection time was associated with secondary task engagement. The comprehensive data analysis indicated that the drivers engaged selectively in secondary tasks in accordance with changes in the demands imposed by driving and roadway situations. The drivers exercised self-regulation by reducing their engagement with secondary activities during more demanding driving situations. This self-regulatory behaviour was represented by the V-shaped relationship between the proportions of time devoted to secondary task engagement across the three intersection stages and the greater willingness of the drivers to engage in such activities when their vehicles were stationary than when the vehicles were moving. The behaviour was also reflected by the diminished willingness of the drivers to engage in secondary tasks when they did not have priority and when they travelled along intersections managed with traffic signs. A particularly important finding is that the elderly drivers were less likely to engage in secondary tasks than the younger drivers.

The results can serve as guidelines for the development of safety measures intended for traffic systems at intersections. They also offer the preliminary information needed to improve driver training/education and awareness programmes on managing distractions and safe driving strategies, especially for novice drivers. Finally, the findings can contribute to the creation of guidelines for classifying intersections in terms of the prevalence and self-regulation of secondary task engagement. These guidelines can be established on the basis of the resultant broadened understanding of who engages in secondary tasks at intersections, when these tasks are executed, what types of tasks drivers occupy themselves with and where such tasks are implemented.

Further research is planned to expand the dataset and further scrutinise the concerted effects of other individual or collective situational factors on drivers' secondary task engagement at intersections. Another initiative under way is the development of a mechanism for using speed data as continuous variables rather than considering them only as binary variables (i.e. moving versus stationary). A limitation worth noting is that no baseline epochs for non-intersection-related behaviours were adopted in this work. Although an examination of upstream and downstream areas of intersections uncover insights, these will not be representative of driving outside intersections.

Acknowledgments

I would like to express appreciation to the UDRIVE project for granting me access to the UDRIVE data. I would also like to thank the UDRIVE local data centre at the University of Leeds team for their guidance and support throughout my study.

References

[1] Regan, M.A., Lee, J.D., Young, K.L.: 'Driver distraction: Theory, effects, and mitigation' (CRC Press, 2009).

[2] Regan, M.A., Hallett, C., Gordon, C.P.: 'Driver distraction and driver inattention: Definition, relationship and taxonomy', Accid. Anal. Prev., 2011, 43, (5), pp. 1771–1781.

[3] Dingus, T.A., Guo, F., Lee, S., et al.: 'Driver crash risk factors and prevalence evaluation using naturalistic driving data'. Proc. Natl. Acad. Sci., 2016, 113, (10), pp. 2636–2641.

[4] TRL., 'Study on Good Practices for Reducing Road Safety Risks Caused by Road User Distractions. Final Report' (European Commission, 2015).

[5] Kircher, K., 'Driver Distraction: A Review of the Literature' (Swedish National Road and Traffic Research Institute, 2007).

[6] Klauer, S.G., Dingus, T.A., Neale V.L., et al., 'The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data' (Virginia Tech Transportation Institute, 2006).

[7] Stutts, J.C., Reinfurt, D.W., Staplin, L., et al., 'The Role of Driver Distraction in Traffic Crashes' (AAA Foundation for Traffic Safety, 2001).

[8] Wang, J.-S., Knipling, R.R., Goodman, M.J.: 'The role of driver inattention in crashes: New statistics from the 1995 Crashworthiness Data System'. Proc. Ann. Adv. Automot. Med., Vancouver,1996, pp. 377-392.

[9] Van Schagen, I., Welsh, R., Backer-Grondahl, A., et al., 'Towards a Large-Scale European Naturalistic Driving Study: Main Findings of PROLOGUE' (SWOV Institute for Road Safety Research, 2011).

[10] Carsten, O., Hibberd, D., Bärgman, J., et al., 'Driver Distraction and Inattention' (EU FP7 Project UDRIVE Consortium, 2017).

[11] Dingus, T.A., Hanowski, R.J., Klauer, S.G.: 'Estimating crash risk', Ergon. Des., 2011, 19, (4), pp. 8–12.

[12] Sayer, J.R.: ‘The effects of secondary tasks on naturalistic driving performance’ (Transportation Research Institute, 2005).
[13] Metz, B., Landau, A., Just, M.: ‘Frequency of secondary tasks in driving - Results from naturalistic driving data’, Saf. Sci., 2014, 68, pp. 195–203.
[14] Charlton, J.L., Catchlove, M., Scully, M., et al.: ‘Older driver distraction: A naturalistic study of behaviour at intersections’, Accid. Anal. Prev., 2013, 58, pp. 271–278.
[15] Stutts, J., Feaganes, J., Rodgman, E., et al., ‘Distractions in Everyday Driving’ (AAA Foundation for Traffic Safety, 2003).
[16] Tivesten, E., Dozza, M.: ‘Driving context influences drivers’ decision to engage in visual–manual phone tasks: Evidence from a naturalistic driving study’, J. Safety Res., 2015, 53, pp. 87–96.
[17] Streubel, T., Rittger, L., Hoffmann, K., et al.: ‘Naturalistic driving behaviour at inner-city intersections’. Proc. ITS World Congr., France, October 2015.
[18] Simon, M., Hermitte, T., Page, Y.: ‘Intersection road accident causation: A European view’. Proc. Int. Tech. Conf. Enhanc. Saf. Veh., 2009, pp. 1-10.
[19] Barnard, Y.F., Utesch, F., van Nes, N., et al.: ‘The study design of UDRIVE: The Naturalistic Driving Study across Europe for cars, trucks and scooters’, Eur. Transport Res. Rev.,2016, 8, (2), pp. 1-10.
[20] Utesch, F., Bärgman, J., Carsten, O., et al., 'Preliminary Analysis Plan' (EU FP7 Project UDRIVE Consortium, 2014).
[21] Eenink, R., Barnard, Y., Baumann, M., et al.: ‘UDRIVE: The European naturalistic driving study’. Proc. Transp. Res. Arena Conf., Paris, France, April 2014, pp. 1-10.
[22] Dingus, T.A., Klauer, S.G., Neale, V.L., et al., ‘The 100-Car Naturalistic Driving Study, Phase II-Results of the 100-Car Field Experiment’ (National Highway Traffic Safety Administration, 2006).
[23] Dotzauer, M., Stemmler, E., Utesch, F., et al., ‘Risk factors, Crash Causation and Everyday Driving’ (EU FP7 Project UDRIVE Consortium, 2017).
[24] Stover, V., ‘Discussion Paper No. 7: Functional Intersection Area’ (Transportation Research Institute, Oregon State University, 1996).
[25] Garber, N.J., Hoel, L.A.: ‘Traffic and highway engineering’ (Cengage Learning, 5th edn. 2014).
[26] DMRB: ‘Design manual for roads and bridges: Geometric design of major/minor priority junctions’ (1995).
[27] DMRB: ‘Design manual for roads and bridges: Highway link design’ (2002).
[28] Sullman, M.J.: ‘An observational study of driver distraction in England’, Transp. Res. Part F Traffic Psychol. Behav., 2012, 15, (3), pp. 272–278.

Stress, fatigue and inattention amongst city bus drivers – an explorative study on real roads within the ADAS&ME project

Christer Ahlström[1*], Anna Anund [1,2], Erik Håkansson Kjellman [1]

[1] Swedish National Road and Transport Research Institute, Linköping, Sweden
[2] Rehabilitation Medicine, Linköping University, Linköping, Sweden
[*]christer.ahlstrom@vti.se

Abstract: Knowledge about normal glance behaviour and typical stress and fatigue levels amongst city bus drivers is very sparse. We therefore conducted an exploratory pre-study with 15 participants during an actual shift in real traffic with passengers. The aim was to gain knowledge about stress, fatigue and glance behaviour during normal operation of a bus, with the subsequent goal to gather data to facilitate upcoming work on driver state detection algorithms targeting the transfer of control between the driver and an autonomous bus. Data collected during the trials include eye tracking, physiology (electrocardiogram, electrooculogram), subjective ratings (sleepiness and stress) and video (driver and road ahead). Lessons learned includes that driving a bus in an urban environment requires frequent sampling of peripheral visual information (why one-camera eye trackers will not work, and why road centre-based distraction detection algorithms will fail) and that physiological data requires personalised algorithms. Regarding the bus drivers' working situation, fatigue and stress levels were generally low, but increased levels of stress and sleepiness existed even in an exploratory experiment like this without any manipulation.

1. Introduction

The goal of doubling travels with public transportation by 2020 requires more efficient operation, and already now working as a bus driver involves much more than just driving the bus. The responsibilities to control where to go, keep track of the timetable, make sure that the bus is on time, oversee and support ticketing, communicate with the operator and interact with the passengers can be overwhelming [1]. On top of that, the bus driver occupation is associated with negative physiological, physical and psychosocial factors related to driver's health [2]. Many of these factors are expected to become more severe in the future and lead to an even more stressful work environment. Despite this awareness, very little research has been conducted to investigate the levels of stress, sleepiness and inattention amongst city bus drivers during an ordinary day at work. In this exploratory study, we aim to investigate city bus drivers' normal fluctuations in stress and fatigue levels, along with visual behaviour, while driving a specific bus route in real traffic with passengers present in the bus.

Driver fatigue in general has received increased attention during recent years and is now considered to be a major contributor to approximately 15 – 30% of all crashes [3-5]. The main cause of driver fatigue is sleepiness due to sleep loss, being awake for too long, and driving during the circadian low [6]. These factors are amplified by obstructive sleep apnoea, a problem shown to be pronounced in the public transport sector [7]. Also, work-related factors such as stress [8, 9] and shift work [10] contribute to driver fatigue. In addition, it is important to consider the type of task [11, 12], as both cognitive underload and overload contribute to the development of fatigue. City bus drivers in particular face work in a stressful and draining work environment on a daily basis, exposing them to the serious risk of driver fatigue [13].

Driver stress is associated with frustration, irritation, negative mood and aggressive driving behaviours such as speeding violations, tailgating, and involvement in minor traffic accidents, in particular under situations of time pressure [14-17]. Social stress, such as personal tragedies or conflicts with significant others, have been estimated to increase the odds of a fatal road accident by a factor of five [18]. These results, from car drivers, are not easily generalised to bus drivers since they must safeguard passenger safety and indeed their own job [2]. That said, city bus driving has been identified as one of the most stressful occupations [19] due to mental and physical exhaustion [20] caused by conflicting pressures to drive safely while maintaining tight schedules in an external environment that the drivers have little control over [21]. Note that stress is a normal physiological response to adapt, cope or adjust with the situation. It is only when driving is interpreted as demanding or dangerous that stress manifests itself as negative, for example in terms of anxiety or worry [22], or as increased heart rate and blood pressure [23].

Driver distraction and inattention poses a significant safety problem both in the personal and public transport sector. In bus driving, inattention and fatigue are considered to be the most common causes of road crashes [24], and crash analyses have particularly highlighted "inattention", "failure to yield" and "not in lane" as causes of fatal city bus accidents [25]. The sources of distraction causing accidents include those that arise from the driving task itself, and those that derive from the additional requirements associated with bus operation, such as passenger and ticketing-related incidents [1]. The most distracting activities are passenger-related and beyond the control of the bus driver [26].

This study is part of the H2020 project ADAS&ME (Adaptive ADAS to support incapacitated drivers mitigate effectively risks through tailor made HMI under automation). ADAS&Me include seven use cases, one of them addressing bus drivers, with the aim to reduce stress and fatigue by automating the docking procedure at the bus stop. This particular scenario has been highlighted by bus drivers to be very stressful since they have to keep track of the passengers, watch out for vulnerable road users outside the bus, and manoeuvre the bus in a smooth and precise manner [27]. By automating the docking procedure, a procedure that requires the driver to be highly attentive, many risky actions related to passenger unloading, pedestrians crossing near bus stops, and driving off from a stop before passengers have time to get seated [24, 28, 29], can be avoided. When transferring the control from the bus to the driver after departing the bus stop, it is necessary that the driver is fit to take back the driving responsibilities. The main focus of the bus use case in ADAS&Me is to design driver monitoring algorithms that ensures that this is the case. If the driver is not ready to take back the control, the bus will initiate a safe stop procedure.

When starting the algorithm design work in ADAS&Me, it was noticed that very little research was available about typical stress and fatigue levels amongst city bus drivers, except for retrospective self-ratings and questionnaires. Neither could we find any information about typical glance behaviour amongst city bus drivers. We therefore found it necessary to carry out an exploratory data collection to get a better understanding of the stress and fatigue levels that can be expected in city bus drivers' during a normal day's work. The aim of this paper is to describe the results from this pre-study. Given the intended applications of algorithm development and automated docking at bus stops, special focus will be devoted to details useful in the upcoming algorithm development work and to driver behaviour in the vicinity of bus stops.

2. Material and methods

2.1. Participants

In total 15 drivers (2 females/13 males, mean age 41±12 years, 11.6 ± 8.2 years of bus driving experience) were involved in the experiment. They had a BMI of 25.9 ± 3.6 and 13 out of 15 drivers reported being satisfied with their working hours. All participants were recruited from Transdev, the local bus operator in the city of Linköping. The bus drivers received a monetary compensation of about 100 Euros.

The study was approved by the regional Ethics committee in Linköping (Dnr 2017/278-31) and all drivers signed an informed consent form.

2.2. Preparations

Sleep diaries and actigraphy (ActiGraph LLC, Pensacola, FL, US) was collected for two days before the experiment day to keep track of the drivers sleep/wake history. The Actigraph was sent to the drivers together with a background questionnaire and the sleep diaries one week before the experiment day. The intention with the sleep diaries and the Actigraphs was to have a possibility to go back and check if potential outliers could be explained by a deviating sleep history.

2.3. Data collection

The bus was equipped with a three-camera head and eye tracking system (Smart Eye Pro ver. 7.0, SmartEye AB, Gothenburg, Sweden), tuned to give high accuracy in the forward gaze direction at a rate of 60 Hz. The eye tracker was connected to a model of the bus, allowing analyses of the objects in the cockpit attracting the driver's gaze. Gaze data points were consequently clustered into glances towards the

Fig. 1. *The cockpit of the bus, including the targets used in the glance analyses.*

following targets (Fig. 1): 1–front window, excluding the road centre area, 2–road centre, defined as a circle with a radius of 8° centred on the modal point of the gaze distribution, 3–left mirror, 4–right mirror, 5–C90 onboard computer, 6–instrument cluster including speedometer, 7–communication radio, 8–FleeTech system, 9–ticket machine, 10–unknown, and 11–lost tracking. The eye tracking system provides a quality indicator in the range from 0–1, based on the contrast between the edge of the iris and the sclera. All samples with gaze quality below 0.2 were set to 'lost tracking' to remove unreliable data. In the current dataset, 34.4±9.9 % of the data were set to 'lost tracking'. The high percentage of lost tracking is likely due to extreme gaze directions outside the cameras' coverage in the present camera setup, especially near bus stops, a large head box (compared to cars), and possibly also larger windows and less shadow, giving rise to more squinting.

Physiological data were acquired with a sampling rate of 256 Hz by a portable digital recording system (Vitaport 2, Temec Instruments BV, the Netherlands). This included an electrooculogram (EOG, measured vertically and horizontally across the eyes) and an electrocardiogram (ECG, lead II). The electrodes used were of the disposable Ag/AgCl type. Electrodermal activity (EDA) was also recorded via a wearable wrist device (Empatica E4, Empatica Inc., Italy).

An observer accompanied the bus driver throughout the experiment. The observer also asked the driver to rate his/her subjective sleepiness level on the Karolinska sleepiness scale (KSS) [30] and stress level on the Stockholm University stress scale (SUS) [31] every fifth minute. These are anchored scales with nine levels, KSS: 1–extremely alert, 3–alert, 5–neither alert nor sleepy, 7–sleepy, no effort to stay awake, and 9–very sleepy, great effort to keep awake, fighting sleep, and SUS: 1–very low stress (very calm and relaxed), 3–low stress (calm and relaxed), 5–neither low nor high stress, 7–high stress (high tension and pressure), 9–high stress (very high tension and pressure).

Kinematics and GPS data were recorded with a data logger that also stored video of the forward view and of the driver (Video VBOX Pro, Racelogic, Buckingham, UK). The data logger was synchronized with the physiological recording system and the eye tracker.

2.4. Design

The design of the experiment was exploratory, and no experimental manipulation of the stress or sleepiness levels of the driver was made. Instead, the drivers' normal fluctuations in stress and sleepiness levels were of interest. This means that except for the electrodes and the measurement equipment, there is no difference between the experiment and an ordinary day at work. The study was run in the medium sized city of Linköping (about 160000 inhabitants). The specific route that was chosen for the test was selected since it has, for a medium sized city, a tight time schedule, since most of the route is on city roads, due to the large number of passengers, and since there are many bus stops along the route. The data collection was done during a normal working day while driving the bus with real passengers. Data from two drivers were collected each day, during the morning shift and during the afternoon shift, respectively. After the shift, the measurement equipment and electrodes were removed, and the driver answered a final questionnaire about his/her experiences during the shift.

The schedule for carrying out the experiment was very tight since extended preparations would interfere with the rest and drive time regulations. In total, we had 20 minutes to inform the participants, attach the electrodes, calibrate the eye tracker and start the data logger. After the trial, we had 4 minutes to power down the system, remove the electrodes, etc.

2.5. Data pre-processing

Blink durations, which is a commonly used measure of sleepiness and fatigue [32], were extracted from the EOG to complement the subjective KSS ratings. Blinks were extracted using an automatic blink detection algorithm based on derivatives and thresholding [33]. To reduce problems with concurrence of eye movements and blinks, the blink duration was calculated at half the amplitude of the upswing and the downswing of each blink and defined as the time elapsed between the two.

Heart rate variability (HRV) metrics and EDA are commonly used measures of driver stress [34]. Heart beats were extracted from the ECG using an automatic detection algorithm based on filter banks [35]. Guided by the meta-analysis by Castaldo et al. [36], three HRV metrics were chosen due to their relation to acute mental stress; the power spectrum density in the HF band (0.15–0.4 Hz), the LF/HF ratio (where LF is the power in the 0.04–0.15 Hz band), and the square root of the mean squared differences between successive heart beats (RMSSD).

The EDA signal was decomposed into a tonic and a phasic component using Ledalab [37], where the phasic component was used as yet another indicator of driver stress.

Table 1: Sleepiness ratings where each value corresponds to the feeling during the past five minutes.

KSS	Frequency	Percentage
1	60	38.0
2	39	24.7
3	41	25.9
4	13	8.2
5	5	3.2
6	0	0.0
7	0	0.0
8	0	0.0
9	0	0.0
Total	158	38.0

The mean and the 90th percentile blink durations as well as the HRV metrics were calculated in a sliding window with 20 seconds overlap and a width of 5 minutes.

2.6. Analysis

Most analyses were based on descriptive statics due to the exploratory nature of the study. *Fatigue indicators* were investigated as a function of time on task, with the expectation that the drivers would become more fatigued over time. *Stress levels* were investigated as a function of how delayed the bus was compared to the time table, and also near bus stops (mean value in the region ± 100 meters from the bus stop) versus in between bus stops. This was analysed with a mixed model analysis of variance (ANOVA) with the fixed factor bus stop versus driving, and the random factors participant and bus stop. *Inattention*, or rather glance behaviour, was analysed as glance frequencies and glance durations throughout the trip. The distribution of glances to the coded glance targets was also analysed near the bus stops.

3. Results

3.1. Sleepiness

On average the bus drivers reported low levels of subjective sleepiness while driving, see Table 1. Two drivers

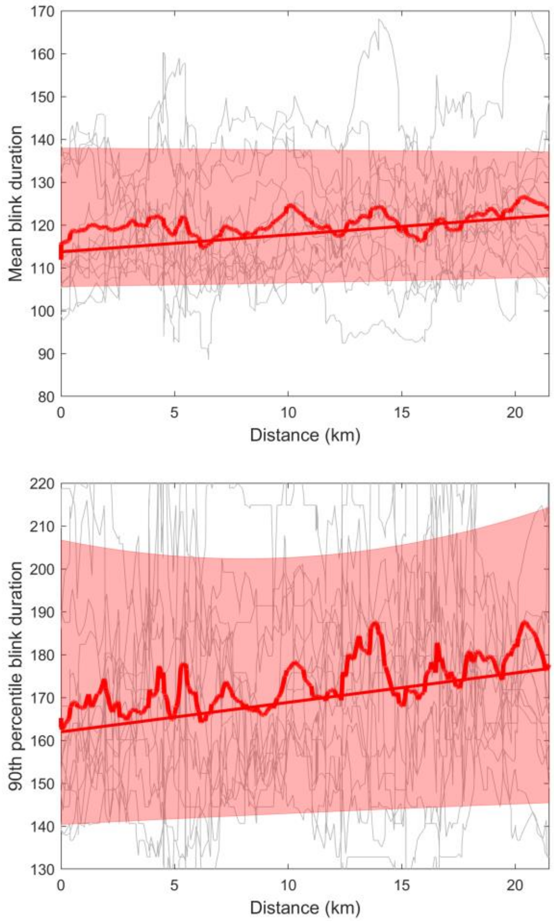

Fig. 2. *Mean and 90th percentile blink durations as a function of distance driven. The grey curves are from individual participants, the red bold lines are the mean across the grey curves and a fitted regression line. The transparent red area is the 10th to 90th quantile regression lines of order 2.*

had mean blink durations exceeding 150 ms in five of the 5-minute segments. This is a clear indication of sleepiness. There was a slight trend towards longer mean blink durations in the end of the drive (median regression line with slope 0.40 and intercept 113.74), see Fig. 2. This trend was stronger when only considering the longest blinks in each 5-minute segment (median regression line with slope 0.69, intercept 161.99), see Fig. 2. In total 5 out of 15 bus drivers reported being sleepy during the data collection. They justified this by: went to bed too late, early morning start, poor sleep the night before, just woke up, too much time waiting at red lights.

3.2. Stress

On average the bus drivers reported low levels of stress while driving, see Table 2. However, some individuals reported high levels at some specific situations even though they were not manipulated. Four out of 15 bus drivers reported high levels of stress in the post-questionnaires. They justified this by: Lots of passengers, problems and misunderstandings, being late, dense traffic, and passengers shouting and talking loudly.

It was hypothesised that higher levels of stress would be reached at bus stops compared to while driving between bus stops. However, the mixed-model ANOVA showed no significant main effects on any of the HRV metrics at the 1 % significance level. There was, however, large individual differences and differences between the various bus stops (Table 3). There was an effect of bus stop versus driving on phasic EDA, but since this finding was not supported by the HRV metrics, this is probably a spurious result, or perhaps an effect of increased sweating caused by manoeuvring the bus near the bus stop.

When comparing HRV metrics versus how delayed the bus was compared to the time table, it was noticed that HF, and to some extent also RMSSD, was reduced with larger delays, see Fig. 3. Above all, the variation in the HF and RMSSD values decreased with the delay, and large delays were characterised by an absence of higher values.

3.3. Inattention

In total one bus driver reported being inattentive. This was justified by: high stress level and misallocated focus on traffic-irrelevant issues.

The glance behaviour data didn't show unexpectadly long glances to any of the in-vehicle systems, table 4 and Fig. 4. The low frequency of glances to the right mirror is probably due to the large head movements, which lead to the loss of visibility of the eye in the used camera setup. Eyes off road glances had a mean duration of 0.7 seconds and a 95th percentile duration of 2.3 seconds, which is comparable to what is typically found in car driving.

Table 2: Stress ratings where each value corresponds to the feeling during the past five minutes.

SUS	Frequency	Percentage
1	59	39.9
2	54	34.0
3	14	8.8
4	12	7.5
5	8	5.0
6	8	5.0
7	3	1.9
8	1	0.6
9	0	0.0
Total	159	100

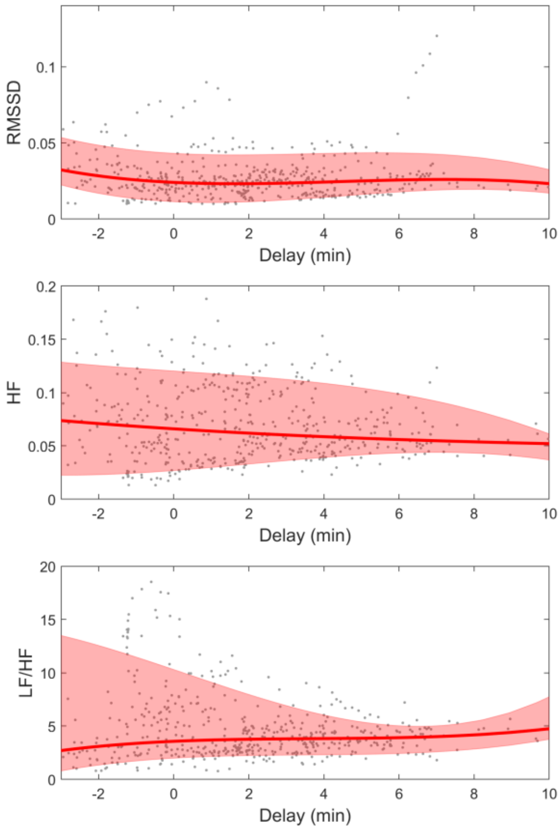

Fig. 3. *HRV metrics plotted as a function of how delayed the bus is compared to the time table. The grey dots are individual HRV metrics per 5-minute segment along the route, the red bold lines and the red shaded areas are the median quantile regression line and the 10^{th} to 90^{th} quantile regression area (order 3).*

Table 4: Glance duration and glance frequency. For 'unknown' and 'no tracking', a "glance" is determined as the "gap" between two known glances.

	Number of glances per km	Mean glance duration (ms)	95th percentile glance duration (ms)
Left mirror	6.0	396	1034
Right mirror	0.6	158	404
C90	3.4	266	875
Cluster	4.3	304	700
Com radio	0.8	282	904
FleeTech	1.5	212	500
Ticket machine	1.7	167	517
Unknown	23.6	347	1084
No tracking	48.4	569	1967
Eyes off road	56.3	731	2317

Glance behaviour while approaching the bus stop showed that when the bus got closer to the bus stop, the drivers looked less in the road centre region and gradually shifted their focus towards the periphery. This visual scanning behaviour is first seen as an increase in the gaze distribution towards the rest of the windscreen (about 5 seconds before arriving at the bus stop), see Fig. 5. When getting even closer to the bus stop, both road centre and windscreen glances are reduced further, and the drivers are only looking in these regions for about 20% of the time. At the same time, the percentage of lost tracking and glances towards unknown glance targets increased. This is probably because the drivers are focusing their attention to vulnerable road users outside the bus and towards the passengers who are lining up to get onboard. The reason why this is coded as no tracking is likely because the gaze direction is outside the coverage of the eye tracking cameras.

Glance behaviour when leaving the bus stop is similar to when approaching the bus stop, but in reversed order. The percentage of road centre and windscreen glances are continuously increasing until they together reach a level of about 75%. The percentage of glances to the left mirror increases during the first five seconds after departure, showing that the drivers are checking for traffic from behind before they depart.

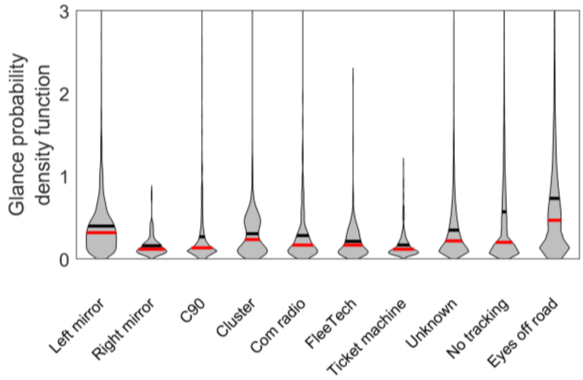

Fig. 4. *Violin plot showing the probability density functions when looking at different glance targets.*

4. Discussion

The bus drivers in this exploratory study of a normal driving shift generally showed low levels of fatigue and stress. This was expected since they followed their ordinary duty roster, without manipulation of stress and fatigue levels. Perhaps the most interesting outcomes from this study concerns methodological aspects and the observed behaviour that couldn't be measured, as outlined below.

Subjective sleepiness ratings based on KSS is a trusted estimate of sleepiness that is as close to a gold standard as we, today, can get. Yet, the bus drivers reported suspiciously low levels of sleepiness during the data collection, much lower levels than is normally seen during alert conditions. The most frequent rating was KSS=1, a condition that essentially means hyperalert, something that only occurs for short lapses. The reasons for this are not known, but it may be because they didn't want to report high levels of stress and sleepiness in front of the passengers. Another reason may be that they hadn't fully understood the rating scale. KSS is a nonlinear scale where the default state should be around 4–5, where lower levels are essentially different levels of alert. Information about the scales were sent to the drivers

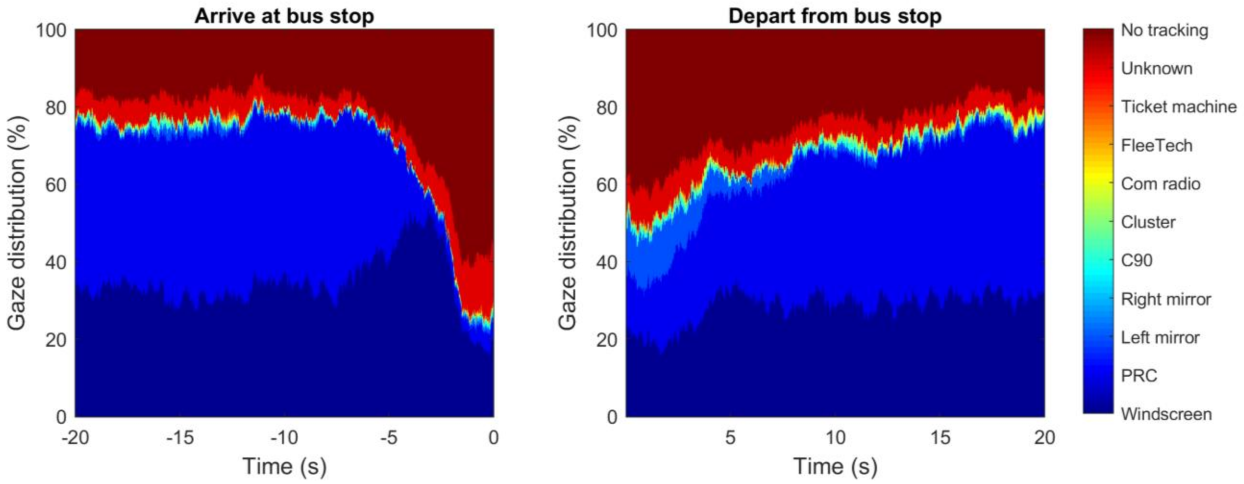

Fig. 5. *Distribution of glances to different targets over time, during the last 20 seconds before stopping at the bus stop and during the first 20 seconds after departing from the bus stop.*

beforehand, but it seems like not all of them had read the instructions thoroughly enough before arrival. However, despite the low self-reported values, some drivers reported high levels of sleepiness in the post-questionnaires, some drivers experienced mean blink durations above 150 ms, and the 90th percentile blink durations showed a clear time on task effect. This supports the explanation that the drivers did not find it comfortable to report their experience while driving.

The subjective stress ratings were also very low, just as the sleepiness ratings, and again, the post-questionnaires revealed that several (4 out of 15) bus drivers had experienced high levels of stress. This was supported by the HRV metrics that indicated decreasing RMSSD, decreasing HF, and slightly increasing LF/HF for larger delays compared to the time table, all indicating elevated stress levels. Again, this indicates the drivers did not find it comfortable to report their experience while driving. This problem with the subjective ratings is difficult to get around. Verbal ratings will always be heard by the passengers, and even with an ethical approval, most bus companies will not allow their drivers to enter the ratings on a tablet or similar device while driving due to company policies.

The most interesting results from the glance behaviour analyses comes from what is inferred from ‘lost tracking’. From Fig. 5, it is painfully obvious how important context is when analysing visual behaviour, especially in complex environments such as in the city. Available real-time driver distraction detection algorithms typically set up a fixed ‘on-road’-region where the driver is supposed to look most of the time. When looking outside this region for too often or for too long, the driver is considered distracted [38]. In Fig. 5, one can see that the ‘on-road’-region must be dynamic, and in this case with the bus stop, this region should represent the bus stop and the vulnerable road users surrounding it rather than the road ahead. The problem is that there is no gaze data available in this direction. In future studies a fourth camera positioned close to the right A-pillar would be beneficial to track right-side glances. Also, ideally, there should not be one but several regions, that change adaptively with the road environment and surrounding road users [39]. To operationalise such an approach, the eye tracking data needs to be fused with environmental sensing. That said, it is clear that automating the docking procedure will help relieve the bus driver in a situation where many targets in multiple locations has to be attended simultaneously.

An observation made during the exploratory analyses was that the self-ratings as well as the physiological measures and the glance data showed variations due to the environment. This indicates that it is important to consider multiple factors simultaneously, and not just multiple physiological indicators, but also external factors (environment, traffic complexity, route, scheduling, passengers etc.) as well as individual aspects (driver traits, health status, family situation etc.), when trying to understand and predict changes in sleepiness, stress and visual attention. Such research has been initiated to get a better picture of the causes of driver sleepiness [40], stress [41] and inattention [39], but mostly on a theoretical level, and not taking the operator's demands into account. There is also very little research on the interaction between multiple simultaneous driver states. Research in this direction is laborious and requires costly experiments, but it may be the only way forward when designing the driver monitoring systems that will (have to be) an integral part of the intermediate steps towards full automation.

5. Conclusions

The results show that even without manipulation there are epochs of sleepiness and stress in some individuals at a normal bus route during daytime.

Countermeasures to make sure this is not the case is most truly helpful for the drivers.

Algorithms that estimate the driver’s state based on physiological data should be personalised.

Driver state detection algorithms, especially for stress and inattention, must take the traffic environment and surrounding road users into account.

6. Acknowledgments

This project was funded by the European Union’s Horizon 2020 research and innovation programme under grant agreement No 688900.

7. References

[1] Salmon, P.M., Young, K.L., and Regan, M.A., 'Distraction ‘on the Buses’: A Novel Framework of Ergonomics Methods for Identifying Sources and Effects of Bus Driver Distraction', *Applied Ergonomics*, 2011, 42, (4), pp. 602-610.

[2] Tse, J.L.M., Flin, R., and Mearns, K., 'Bus Driver Well-Being Review: 50 Years of Research', *Transportation Research Part F: Psychology and Behaviour*, 2006, 9, (2), pp. 89-114.

[3] Herman, J., Ameratunga, S., Wainiqolo, I., Kafoa, B., Robinson, E., McCaig, E., and Jackson, R., Prevalence of Sleepiness While Driving Four-Wheel Motor Vehicles in Fiji: A Population-Based Survey (Trip 9)', *Injury prevention : journal of the International Society for Child and Adolescent Injury Prevention*, (2013, 2013/01/17 edn.)

[4] Williamson, A., Lombardi, D.A., Folkard, S., Stutts, J., Courtney, T.K., and Connor, J.L., 'The Link between Fatigue and Safety', *Accident Analysis and Prevention*, 2011, 43, (2), pp. 498-515.

[5] Connor, J., Norton, R., Ameratunga, S., Robinson, E., Civil, I., Dunn, R., Bailey, J., and Jackson, R., 'Driver Sleepiness and Risk of Serious Injury to Car Occupants: Population Based Case Control Study', *Br. Med. J.*, 2002, 324 p. 1125.

[6] Åkerstedt, T., Connor, J., Gray, A., and Kecklund, G., 'Predicting Road Crashes from a Mathematical Model of Alertness Regulation - the Sleep/Wake Predictor', *Accid Anal Prev*, 2008, 40, (4), pp. 1480-1485.

[7] Firestone, R.T., Mihaere, K., and Gander, P.H., 'Obstructive Sleep Apnoea among Professional Taxi Drivers: A Pilot Study', *Accid Anal Prev*, 2009, 41, (3), pp. 552-556.

[8] Chen, C.-F. and Kao, Y.-L., 'The Connection between the Hassles-Burnout Relationship, as Moderated by Coping, and Aberrant Behaviors and Health Problems among Bus Drivers', *Accid Anal Prev*, 2013, 53, (Journal Article), pp. 105-111.

[9] Karasek, R. and Theorell, T., Healthy Work: Stress, Productivity, and Thew Receonstruction of Working Life.', in Books, B. (ed.), (1990)

[10] Philip, P. and Akerstedt, T., 'Transport and Industrial Safety, How Are They Affected by Sleepiness and Sleep Restriction? ', *Sleep Medicine Reviews*, 2006, 10, pp. 347-356.
[11] Sallinen, M., Härmä, M., and Akila, R., 'The Effects of Sleep Debt and Monotonous Work on Sleepiness and Performance During a 12-H Dayshift', *J Sleep Res*, 2004, 13, pp. 285-294.
[12] Thiffault, P. and Bergeron, J., 'Monotony of Road Environment and Driver Fatigue: A Simulator Study', *Accident Analysis and Prevention*, 2003, 35, (3), pp. 381-391.
[13] Biggs, H., Dingsdag, D., and Stenson, N., 'Fatigue Factors Affecting Metropolitan Bus Drivers: A Qualitative Investigation', *Work*, 2009, 32, (1), pp. 5-10.
[14] Gulian, E., Glendon, A., Matthews, G., Davies, D., and Debney, L., 'The Stress of Driving: A Diary Study', *Work & Stress*, 1990, 4, (1), pp. 7-16.
[15] Hennessy, D.A. and Wiesenthal, D.L., 'Traffic Congestion, Driver Stress, and Driver Aggression', *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, 1999, 25, (6), pp. 409-423.
[16] Hennessy, D.A., Wiesenthal, D.L., and Kohn, P.M., 'The Influence of Traffic Congestion, Daily Hassles, and Trait Stress Susceptibility on State Driver Stress: An Interactive Perspective', *Journal of Applied Biobehavioral Research*, 2000, 5, (2), pp. 162-179.
[17] Matthews, G., Dorn, L., and Glendon, A.I., 'Personality Correlates of Driver Stress', *Personality and Individual Differences*, 1991, 12, (6), pp. 535-549.
[18] Brenner, B. and Selzer, M.L., 'Risk of Causing a Fatal Accident Associated with Alcoholism, Psychopa - Thology, and Stress: Further Analysis of Previous Data', *Systems Research and Behavioral Science*, 1969, 14, (6), pp. 490-495.
[19] Evans, G.W., Johansson, G., and Rydstedt, L., 'Hassles on the Job: A Study of a Job Intervention with Urban Bus Drivers', *Journal of Organizational Behavior*, 1999, pp. 199-208.
[20] Karasek Jr, R.A., 'Job Demands, Job Decision Latitude, and Mental Strain: Implications for Job Redesign', *Administrative science quarterly*, 1979, pp. 285-308.
[21] Gardell, B., Aronsson, G., and Barklöf, K., The Working Environment for Local Public Transport Personnel', (Swedish Work Environment Fund, 1982)
[22] Gulian, E., Debney, L.M., Glendon, A.I., Davies, D.R., and Matthews, G., Coping with Driver Stress', *Stress and Tension Control 3*, (Springer, 1989)
[23] Robertson, S., 'Stress Related Driving: An Exploratory Study', 1988.
[24] Cafiso, S., Di Graziano, A., and Pappalardo, G., 'Using the Delphi Method to Evaluate Opinions of Public Transport Managers on Bus Safety', *Safety science*, 2013, 57, pp. 254-263.
[25] Blower, D., Green, P.E., and Matteson, A., 'Bus Operator Types and Driver Factors in Fatal Bus Crashes: Results from the Buses Involved in Fatal Accidents Survey', 2008.
[26] D’Souza, K. and Maheshwari, S., 'Improving Performance of Public Transit Buses by Minimizing Driver Distraction', *WIT Transactions on The Built Environment*, 2012, 128, pp. 281-293.
[27] Willstrand, T., Anund, A., Strand, N., Nikolaou, S., Gemou, M., and Faller, F., Adas&Me Deliverable 1.2 – Driver/Rider Models, Use Cases and Implementation Scenarios', (2017)
[28] Greiner, B.A., Krause, N., Ragland, D.R., and Fisher, J.M., 'Objective Stress Factors, Accidents, and Absenteeism in Transit Operators: A Theoretical Framework and Empirical Evidence', *Journal of occupational health psychology*, 1998, 3, (2), p. 130.
[29] Kompier, M.A., *Bus Drivers: Occupational Stress and Stress Prevention*, (International Labour Office Geneva, 1996)
[30] Åkerstedt, T. and Gillberg, M., 'Subjective and Objective Sleepiness in the Active Individual', *Int J Neurosi*, 1990, 52, pp. 29 - 37.
[31] Dahlgren, A., Kecklund, G., and Akerstedt, T., 'Different Levels of Work-Related Stress and the Effects on Sleep, Fatigue and Cortisol', *Scand J Work Environ Health*, 2005, 31, (4), pp. 277-285.
[32] Schleicher, R., Galley, N., Briest, S., and Galley, L., 'Blinks and Saccades as Indicators of Fatigue in Sleepiness Warnings: Looking Tired?', *Ergonomics*, 2008, 51, (7), pp. 982-1010.
[33] Jammes, B., Sharabty, H., and Esteve, D., 'Automatic Eog Analysis: A First Step toward Automatic Drowsiness Scoring During Wake-Sleep Transitions', *Somnologie-schlafforschung und Schlafmedizin*, 2008, 12, (3), pp. 227-232.
[34] Healey, J.A. and Picard, R.W., 'Detecting Stress During Real-World Driving Tasks Using Physiological Sensors', *IEEE Transactions on intelligent transportation systems*, 2005, 6, (2), pp. 156-166.
[35] Afonso, V.X., Tompkins, W.J., Nguyen, T.Q., and Luo, S., 'Ecg Beat Detection Using Filter Banks', *IEEE transactions on biomedical engineering*, 1999, 46, (2), pp. 192-202.
[36] Castaldo, R., Melillo, P., Bracale, U., Caserta, M., Triassi, M., and Pecchia, L., 'Acute Mental Stress Assessment Via Short Term Hrv Analysis in Healthy Adults: A Systematic Review with Meta-Analysis', *Biomedical Signal Processing and Control*, 2015, 18, pp. 370-377.
[37] Benedek, M. and Kaernbach, C., 'A Continuous Measure of Phasic Electrodermal Activity', *Journal of Neuroscience Methods*, 2010, 190, (1), pp. 80-91.
[38] Ahlström, C. and Kircher, K., 'Review of Real-Time Visual Driver Distraction Detection Algorithms', in, *Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research*, (ACM, 2010)
[39] Kircher, K. and Ahlstrom, C., 'Minimum Required Attention: A Human-Centered Approach to Driver Inattention', *Human factors*, 2017, 59, (3), pp. 471-484.
[40] Van Dongen, H., Vitellaro, K., and Dinges, D., 'Individual Differences in Adult Human Sleep and Wakefulness: Leitmotif for a Research Agenda', *SLEEP*, 2005, 28, (4), pp. 479-496.
[41] Matthews, G., 'Towards a Transactional Ergonomics for Driver Stress and Fatigue', *Theoretical Issues in Ergonomics Science*, 2002, 3, (2), pp. 195-211.

What were they thinking? Subjective experiences associated with automation expectation mismatch

Pär Gustavsson[1*], Trent W. Victor [1], Joel Johansson[1], Emma Tivesten[1], Regina Johansson[1] and Mikael Ljung Aust[1]

[1] Volvo Cars Safety Centre, Volvo Cars, Gothenburg, Sweden
[*]par.gustavsson@volvocars.com

Abstract: The aim of this paper is to gain insight into factors that affect drivers' mental processes and responses to a critical event while driving with supervised automation. Seventy-six drivers participated in a test track experiment ending with a conflict event that required active driver intervention to avoid a crash. About a third of the drivers crashed, despite being provided with instructions on system limitations as well as supervision reminders. Analysis of questionnaire and interview data from the drivers showed that crash outcomes could not be explained by factors often brought up as concerns when discussing supervised automation, such as sleepiness and inattention. Instead, the drivers who crashed did so due to expectation mismatch. This in turn seems to stem from learned trust. Crashers reported higher trust in the automation than non-crashers. Crashers also saw the conflict object but believed the vehicle would be able to resolve the situation on its own. For some drivers, 30 minutes of driving with highly reliable supervised automation thus seems to provide sufficient grounds for developing incorrect expectations on automated function capabilities. These conclusions are relevant to human-automation interaction in general, and for development of driver-state-adaptive supervised automation and advanced driver-assistance systems in particular.

1. Introduction

Development of successful automated driving will depend on recognizing and supporting the two new driver roles that come with driving automation – the delegated and the shared driving role, or unsupervised and supervised automated driving respectively [1].

In unsupervised automation, the driver delegates full control and responsibility to the vehicle, to be free to do something else (e.g. work, watch a film, or even sleep). This requires a vehicle designed for complete support and crash avoidance in all conflict situations (see e.g. [2] and [3]).

Supervised automation, on the other hand, only partly supports the driving task (e.g. headway control and some degree of steering assistance), and the driver is still required to supervise the driving and intervene at sensing or actuation limits (e.g. conflict situations). The driver is thus not free to disengage from the driving task.

In this context it is important to note that meta-analyses [4] has showed that there exists a general relationship between degree of automation and reduction in human performance (such as complacency, skill degradation, and loss of situation awareness). For instance, it has been found that while increased automation improves routine task performance, operators show difficulty troubleshooting and recovering when something unexpected happens [4].

This human performance reduction is largely attributed to operators' tendency to reduce their monitoring of highly reliable automation because of its ability to function properly for an extended period of time (e.g. [5] and [6]). It is simply difficult for humans to monitor automation, or be out of the loop for some time, and then suddenly solve critical issues [7].

Also, note that attention and understanding are often implicitly mixed together in descriptions of monitoring [8] [9], i.e. many assume that as long as a conflict object is perceived, it will be adequately acted upon. However, looking at an object or a road segment does not necessarily mean that cognitive control (top down selection processes) becomes engaged and actions are executed [10] [11].

The difficulty to monitor automation is often referred to as an Irony of Automation [12]. In short, as automation becomes more reliable during routine driving and the operational design domain expands (e.g. more situations, speeds, and road types), drivers may develop misconceptions that the automation can handle all safety conflict situations, leading to driver disengagement and performance reduction.

For road traffic, this means that the better the automation, the less attention drivers will pay to traffic and the system, and the less capable they will be to resume control [1]. Of particular concern are first failure effects. These are circumstances where the operator encounters perfect automation for some period of time, and then "complacency" or overtrust in automation is reflected by a very poor response when the automation fails [13] [14] [15]. In a simulator study it has also been found that takeover requests (from automation to drivers) were perceived as automation failures and temporarily reduced drivers' automation trust, but that trust still was higher in the end of the drive compared to the start. A possible explanation to this is that the takeover request illustrated that the system was not perfectly reliable, and in the long run might have helped drivers to understand the system, thereby increasing trust [16].

Several concepts have been proposed for understanding the human performance degradation associated with increased automation reliability. For example, a meta-review argued that trust formation is a key concept for understanding human relationships with automation [17].

This analysis describes three layers of variability in human-automation trust: dispositional, situational and learned trust. Of particular interest here is learned trust, which can be further subdivided in two parts: initial learned trust (i.e. trust prior to interacting with the system) and dynamic learned trust (i.e. trust built during interaction). In the context of supervised automation, the former represents what you know about a vehicle automation function before starting to drive, while the latter represents what happens with your trust when you use that function while driving.

It has also been suggested that automation which fails to adapt, i.e. does not change to match the needs of the current situation, may be more susceptible to operator performance degradation [18] [19]. However, while the concept of adaptive automation seems to be viewed as positive, practical applications proving its value are scarce [17]. In the context of supervised automation, one possible implementation of adaptivity would be to let the function remind the supervising operator about monitoring and response readiness responsibilities, should the operator show tendencies to fail at these.

Another concept that is discussed is that of automation transparency. Transparency refers to the extent to which the inner workings or logic of an automated system are transparent to the operator [20]. The general idea is that transparent systems which provide accurate and useful feedback can reduce automation misuse or disuse [17].

However, it is less clear how to apply this in supervised automated driving. While transparency may be intrinsic to establishing initial learned trust (i.e. telling the future function users about function limitations), it is not clear what would constitute transparency in the dynamic phase. Ideas such as displaying which traffic elements the vehicle is actually tracking to the driver [1] to make it easy to spot tracking errors are abundant in online commentaries, but so far, the authors of the current paper are not aware of research where this has been further studied empirically.

The data used in the present paper comes from three experiments examining driver intervention response to conflicts after driving with highly reliable supervised automation following a lead vehicle on a test track [21]. In all experiments a conflict occurred after 30 minutes wherein the lead vehicle cut out of lane to reveal a conflict object in the form of either a stationary car or a garbage bag. Data from the second and third experiment reported in [21] are included in this paper.

In the first (baseline) experiment the test vehicle automatically braked and avoided the conflict. In this experiment, drivers displayed both extreme visual distraction and sleepiness, and few tried to intervene in the conflict. This raised the question of whether these drivers would have avoided the conflict, had the vehicle not intervened on its own.

In the second and third experiment, participants were given more detailed instructions on system limitations and driver responsibilities, Supervision reminders (Attention Reminder and Integrated Attention and Hands on Wheel reminder) were implemented, and the drivers needed to intervene to avoid a crash in the conflict event. Differences between experiments and conditions are described in Table 1 in section 2.1.

Supervision reminders effectively maintained eyes-on-path and hands-on-wheel. However, neither these reminders nor explicit instructions on system limitations and supervision responsibilities prevented 28% (21/76 drivers) from crashing with their eyes on the conflict object. The crash rates were similar across experiments and test conditions [21]. These results highlight the important role of expectation mismatches, showing that a key component of driver engagement is cognitive (understanding the need for action), rather than purely visual (looking at the threat), or having hands-on-wheel [21].

The aim of this paper is to gain insights into the cognitive mechanisms underlying the drivers' expectation mismatch by analysing questionnaire and interview data from the participants in the second and third experiment. Specifically, the perceived relevance of Supervision reminders and which factors that affect drivers' mental processes and actions in a critical event (what the drivers expected themselves to do and the car to do) is explored.

2. Method

This section covers the general test setup, the differences between experimental conditions, data collection, data processing, coding and review, and data analysis. A more detailed method description is provided in [21].

2.1. General method

The same general methodology was used in the two experiments included in this paper (experiment 2 and 3). The main differences between the experiments were the level of instruction detail, supervision reminder type, and conflict scenario type. Key differences between the experiments and conditions are summarized in table 1.

Table 1 *Overview of key differences between experiments and conditions. Supervision Reminder type is either AR (Attention Reminder) or AR&HoW (Integrated Attention and Hands on Wheel reminder).*

Experiment and condition	N	Level of instruction detail	Supervision reminder type	Conflict scenario type
1a	*15*	*Low*	*None*	*Stationary car fully in lane*
1b	*15*	*Low*	*None*	*Stationary car partially in lane*
2	16	Medium	AR	Drift & Garbage bag in lane
3a	15	High	AR	Stationary car partially in lane
3b	15	High	AR	Garbage bag in lane
3c	15	High	AR&HoW	Garbage bag in lane
3d	15	High	AR&HoW	Stationary car partially in lane

2.1.1. Participants:

All 76 participants were Volvo Cars employees. The experiments were set up to achieve a between-group design. The selected participants could not be involved in driving automation development, could not work as test drivers, had not participated in similar studies before, and had a minimum driving experience of at least 5000 km during the previous year.

In experiment 2, 16 participants were included in the final sample, 2 females and 14 males. Ages spanned from 27-66 years (M = 45.9, SD = 12.0) with driving experience spanning from 6-49 years (M = 26.9, SD = 13.7).

In experiment 3, 60 test participants were included in the final sample, 18 female and 42 male. Ages spanned from 26-65 years (M = 45.2, SD = 9.6) and driving experience from 1-47 years (M = 25.3, SD = 10.4).

2.1.2. Materials, procedure and scenario design:

On arrival, the participants received general information about the test and were asked to read through written participant information as well as sign an informed consent form. The participants were also asked to fill in a pre-drive questionnaire in order to provide driver background information. The stated purpose of the study was to evaluate driver experiences during automated driving.

Next, the participants were introduced to the test vehicle (TV), a Volvo XC90 (MY2016). The original XC90 Driver Information Module (DIM) was modified to display a customized supervised automation HMI which also could present attention reminders to participants according to predefined thresholds for visual inattention, a similar algorithm to [22], and the MDD algorithm [23]. The TV was equipped with special, test-unique software which had self-driving capability to precisely follow the road, maintain speed, and keep a constant headway with highly-reliable driving performance behind a robot-controlled XC90 lead vehicle (LV) on the AstaZero rural road test track.

While sitting in the driver's seat, the participants received further verbal information about the test and the vehicle, along with an introduction to the Karolinska Sleepiness Scale (KSS) [24]. The participants reported their sleepiness level (KSS 1-9) before the drive started and once every lap on the test track (approximately every 6 minutes). The participants were instructed to supervise the car throughout the drive and were also told that they could override the automation by steering or braking at any time. Two test leaders rode along in the backseat of the TV; one who administered the KSS scale and monitored a video stream of the driver to trigger supervision reminders according to pre-defined rules for eyes-off-path and hands-off-wheel, and another who acted as (back-up) safety driver. There was no conversation with the test participants during the drive, except when asking for KSS scores.

The TV followed behind the LV which kept a speed of 70 kph, except through some curves where speed was lowered by the LV to about 50 kph. The same (pre-recorded) LV path and velocity was used for all participants. After 30 minutes (five laps), the test vehicle encountered a conflict object placed in the driving lane; either a stuffed garbage bag (figure 1a) or the ADAC Advanced Emergency Braking System Stationary Target (stationary car) (figure 1b). The conflict object was positioned so that the participants could see it when passing through a curve and crest just prior to the event, 14.0 s before reaching the conflict object. The conflict object then became obscured again by the LV when the road straightened out. About 20 meters from the conflict object, the LV did a cut-out (an evasive steering manoeuvre around the object) revealing the conflict object in full to the participants, about three seconds before reaching the object. The TV did not brake or warn the drivers in any way, and the DIM displayed the same HMI throughout the whole drive.

After the conflict, the participants were asked to stop the car and fill in a post-drive questionnaire, which also served as a basis for a semi-structured interview. All interviews were recorded. After the interview, the full purpose of the study was disclosed.

Figure 1a - *Garbage bag used in experiment 2, 3b and 3c.*

Figure 1b – *Stationary car used in experiment 3a and 3d.*

2.2. Experiment 2

Experiment 2 (E2) examined if driver intervention performance is associated with first failure effects [13]. This was done by exposing the participants to a drift out of lane event after 15 minutes, and a conflict situation with a garbage bag inside the lane after 30 minutes. In the drift event, the vehicle drifted over into the left-adjacent lane and returned to the right lane after some time if the driver did not intervene (between 8 and 18 s for the participants who did not steer back).

In experiment 2, participants were given a medium level of instruction – written instructions that emphasized the driver's role as supervisor, the limitations of the vehicle, and the driver's responsibility for the safety of the vehicle even when the automation was engaged. The instruction also stated that the drivers needed to apply more force to override the steering when the automation was engaged compared to what is needed for normal steering in manual driving. A key excerpt from these instructions:

"The car you will drive is a so called Supervised automated drive car which means that the car itself, under certain circumstances and on chosen road stretches, can control steering and adapt speed and distance. Due to limitations in the car's sensor platform the driver can't yet engage in non-

driving activities, and you are instead expected to supervise the drive at all times, as you would in manual driving."

Participants received attention reminders in the DIM (warning messages in the instrument cluster behind the steering wheel) if they were visually inattentive (determined from patterns of off-path eye glances). Two levels of attention reminders were used. Both levels of reminders were presented for a maximum of 7.0 s without any notification sound. If participants looked back on the road for at least 2.0 s after the reminder or were judged to be more attentive, the system reset and the reminder disappeared from the display.

Level one reminders (figure 2) were issued if a single off-path glance longer than 3.4 s was detected, or if the driver had been looking predominantly off-path for a period of 12.0 s (total glance duration history).

Level two reminders were issued for single off-path glances longer than 7.0 s, if the driver's attention did not return to the road after having been issued a level one reminder, for eye-closures longer than 3.0 s, or if they received a new level one reminder within 10.0 s of a level one or two reminder. The only visual design difference for the level two reminder was a red icon. In addition, the level two reminder was combined with a soft deceleration of the test vehicle.

Figure 2 – *Level one Attention Reminder in E2.*

2.3. *Experiment 3*

Experiment 3 (E3) examined if more detailed instructions, updated Attention Reminder rules, and adding an Integrated Hands on Wheel and Attention Reminder improved driver intervention performance. The specific TV used in E3 had slightly improved lane keeping capability and slightly more steering wheel resistance when the automation was engaged than the TV in E2. All participants in E3 were instructed to override the steering approximately one minute into the drive to experience the steering wheel resistance and to minimise potential effects of this on crash outcomes.

The participants were exposed to a conflict situation after 30 minutes with either a stationary car partially inside the lane (condition 3a & 3d) or a garbage bag inside the lane (condition 3b & 3c).

In experiment 3, participants were given a high level of instruction by attending a 30 minute classroom training prior to their drive. The training covered these areas:

Driver responsibilities. The driver is responsible, should monitor, supervise and intervene whenever needed. The driver needs to be active and attentive at all times and supervise the traffic so that the car is driven in a safe manner for passengers in the vehicle and surrounding traffic. Sensors and cameras judge the driver's ability to actively supervise the automation and traffic and detects if the driver has their hands on the steering wheel or if the driver looks on the road. Drivers will get notifications after periods of inattentiveness or inactivity and the system will deactivate after a longer period of inactivity.

System limitations. Objects and obstacles in the traffic environment, such as potholes in the roadway, high curbs, and objects on road are not detected. Obstacles can also be falsely detected as lane markings and thus pose a risk that the car will collide with these obstacles. Cameras and sensors have a limited field of view. Indistinct lane markings might lead to erroneous steering by the automation. Other limitations may occur with road design (e.g. roadworks), oncoming vehicles, pedestrians, and animals. There are restrictions in the steering, braking, and acceleration force that can be applied by the system.

Instruction videos & risk scenarios. Videos of risk scenarios were shown, including a video showing when a car starts to depart from the roadway and the driver needs to steer back in lane and then let the function resume control. A risk scenario where the function does not detect obstacles in the roadway was explained, in which the driver needs to brake and/or steer away from the obstacle and after that let the function resume control.

All participants received attention reminders if they were visually inattentive. The Attention Reminder was updated based on feedback received in experiment 2. Participants in condition 3a and 3b were not required to have their hands on the steering wheel as long as they stayed visually attentive. Three levels of attention reminders were used.

Level one reminders (figure 3, left) were issued in the DIM if participants had been looking predominantly off-path during a period of > 17.0 s (total glance duration history).

Level two reminders used the same message but added a sound and were triggered either by a 3.4 s off-path glance, an eye closure longer than 3.0 s, or if they received a new reminder within 10.0 s of a level one or two reminder.

Level three reminders were issued as a text message *"Autopilot deactivated – Driver inattention"* with a hands-on-wheel icon and a more urgent sound if a 15.0 s glance off path was detected, or a 15.0 s eye closure, or if they were glancing more than 75% off path in a period of 20.0 s (glance history), or if they were to receive a third level two reminder within 15.0 s.

Participants in condition 3c and 3d were required to always keep their hands on the steering wheel and received Hands on Wheel (HoW) reminders if they failed to do so. Thus, drivers in these conditions could experience both attention reminders and Hands on Wheel reminders at different periods during the same trip.

Two levels of Hands on Wheel reminders (figure 3, right) were issued in the DIM.

Level one HoW reminders were issued if hands were off the steering wheel for more than 5.0 s.

Level two HoW reminders used the same message and icon but added a sound and were issued if hands were off the steering wheel for more than 10.0 s.

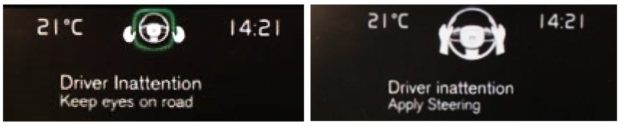

Figure 3 – *Supervision Reminder messages in E3. Left: Attention Reminder message. Right: Hands on Wheel reminder message.*

2.4. Data processing, coding and review

Video of the conflict event for all participants was reviewed in order to determine crash outcome [21]. Any contact between TV and the conflict object was classified as a crash.

All questionnaire responses and rating scales were compiled into a data set. The interview recordings (ranging between 8 and 36 minutes long per participant) were transcribed. The transcriptions were transformed into open codes separated by interview question and participant. In addition, the free text responses in the questionnaires were coded and added to the transcription codes. Themes were created by processing the codes related to one specific interview question or created by processing codes from different parts of the interview, according to content analysis methodology [25]. The themes and included codes were reviewed among the authors in order to reach consensus on the themes.

The process from transcription to themes is presented by the following example:

Questionnaire item: Did you perceive that the object was in the lane before the lead vehicle steered away (y/n)? If yes, when?

Questionnaire item response: No.

Transcription: "No I did not. It was not until the lead vehicle steered away that I understood that there was something there"

Open coding: Did not perceive object until LV manoeuvre

Final theme: I perceived the object late

The categorization of driver actions was based on responses to the interview question "*Did you intervene in the situation? If yes, how and why? If no, why?"*. Responses indicating that the drivers intervened without delay were categorized as *I intervened*. Responses that explicitly described interventions as late or delayed waiting for intervention by the automation were categorized as *I intervened late*. Furthermore, if a participant stated that they realized the need to intervene late and intervened, they were categorized as intervening late as well. The category *I intervened too late* includes statements of acting too late to be able to avoid a crash. If a participant answered no to the interview question or stated that she only put her hands on the steering wheel, it was categorized as *I did not intervene*.

The categorization of realization of the need to act was based on responses to the interview question *"Did you realize that you needed to intervene to avoid a crash? If yes, when did you realize that?"*. The category *I realized the need* includes statements that the participants realized the need to intervene before or at the time of the LV cut-out manoeuvre. *I realized the need late* includes explicit statements that they realized it late, after the LV manoeuvre, or when they were close to the object. The category *I realized the need too late* includes explicit statements that the realization was too late to be able to act and avoid a crash. If a participant responded no to the interview question or stated that they realized the need to intervene after crashing it was categorized as *I did not realize the need*.

The categorization of conflict object perception was based on responses to the interview question "*Did you perceive that the object was in the lane before the lead vehicle steered away? If yes, when?"*. The category *I perceived the object early* means before the LV cut-out manoeuvre (14 s before the conflict point), while *I perceived the object late* includes perception at the time of the LV manoeuvre or later. Some participants also responded that they perceived the object before the LV manoeuvre but were not sure if it was positioned in or outside the lane, creating the category *I perceived the object early, uncertain if in lane.*

The categorization of expectations on the automation was based on an analysis of the responses to several different interview questions, mainly *"Did you intervene in the situation? If yes, how and why? If no, why?"*, *"Did you realize that you needed to intervene to avoid a crash? If yes, when did you realize that?"* and *"To what extent did you trust the automated vehicle to be able to handle the situation?"*. Data from drivers that expressed that they were expecting an intervention from the automation were categorized as *I expected an intervention*. Drivers that expressed uncertainties regarding the automation's ability to intervene were categorized as *I was uncertain about an intervention*. The drivers that stated that they did not expect the automation to intervene were categorized as *I did not expect an intervention.*

The categorization of trust in the automation was based on ratings on a Likert scale (1 not at all, 7 completely) to the question *"To what extent did you trust the automated vehicle to be able to handle the situation?"*. The category *I had high trust* includes ratings 5-7, the category *I was neutral* includes rating 4, and the category *I had low trust* includes ratings 1-3.

2.5. Data analysis and visualizations

The participants' experimental condition, crash outcome, ratings and themes were compiled into a Microsoft Excel sheet. This enabled filtering of different variables in order to analyse the data using a combination of qualitative and quantitative methods.

The drivers' actions and mental processes related to the conflict situation were visualized in a format inspired by the graphical representation used in the DREAM methodology [26].

3. Results

This section presents the results from the analysis of the questionnaire and interview responses.

3.1. Attention Reminder relevance

To study the participants' experiences of attention reminders, the participants who were aware of receiving reminders in experiment 2 and 3 rated the relevance of them on a scale between 1 (not at all relevant) and 7 (very relevant). There was an increase in rated relevance of attention reminders from E2 ($M=4.17$, $SD=1.70$, $N=12$) to E3 ($M=5.75$, $SD=1.25$, $N=48$); $t(14)=-3.03$, $p=0.009$. The increase in relevance of the reminders in E3 was independent of whether hands were on wheel or not.

The drivers were also asked to explain their relevance rating of attention reminders. In E2 only 17% (2/12) of the participants expressed that the reminders were warranted or relevant compared to 77% (37/48) in E3. Further, 67% (8/12) of the participants in E2 found the system to be too sensitive or remind too frequently, which only 23% (11/48) expressed in E3.

Thus, the rated relevance of attention reminders increased on average from a neutral level in experiment 2 to a high level in experiment 3.

3.2. Hands on Wheel reminder relevance

The participants who were aware of receiving Hands on Wheel reminders in condition 3c and 3d rated the relevance of the reminders on a scale between 1 (not at all relevant) and 7 (very relevant). The Hands on Wheel reminders were on average rated as highly relevant (M=5.86, SD=1.07, N=7).

3.3. General themes

This section presents some general reoccurring themes in the data that are relevant to the supervising role drivers were instructed to take in these experiments.

3.3.1. Sleepiness:

An issue mentioned by 32% (n=24/76) of the participants was the risk of becoming sleepy while supervising, especially during longer drives. Only two participants stated in the interview that they were feeling less or equally sleepy compared to manual driving. Some example quotes that highlight problems with sleepiness are (words in [] represent test leader utterances while words in () are added for clarification):

"I felt that I got drowsy after a while… [if we had driven for thirty more minutes, do you think you would have stayed awake?] I would probably have ended up on an 8 on the scale (KSS), I definitely think so… I would have stayed awake but I probably would have needed to fight it a bit…"

"The more you trust the car, the sleepier you get"

"I was very tense the whole drive… it was comfortable to ride along but it felt more laborious than driving yourself since I know that I must control but not drive… know that I must have control but at the same time not have control because it is the car that drives. It is probably a matter of habit but I think that it took a lot of energy and effort since I also do not know how the car will react. When I drive myself I have complete control and I am aware that it is only me that has complete control. When I did not have that it felt like I strained myself very much more and became very tired in the end because it takes a lot of energy"

However, sleepiness does not seem to be a factor that explains why some drivers crashed. On average, the crashers in experiment 3 rated themselves as more tired on the KSS scale for the last lap (M=5.06, SD=1.18, N=16) than non-crashers (M=4.30, SD=1.55, N=44); $t(34)=2.04$, $p=0.049$, but the difference is small and in the middle of the scale. Also, there were no clear differences observed in E2, and only one participant (in E3) reported extreme sleepiness (KSS≥8) during the last lap.

3.3.2. Attentiveness:

Another common concern expressed by 24% (n=18/76) of the participants was problems staying attentive while supervising the drive. Some examples of this are:

"I would like something to keep me more active during the drive, because it is hard to keep your eyes on the road this long"

"It is hard to stay focused, your mind wanders"

"I thought it was hard to be focused but still not have an active role. Even if you sat and looked on the road I experienced that I had lots of other thoughts inside my head"

"When you feel safer you lose focus more and more. You lose the ability to concentrate".

However, in the current study, there was no clear mapping between concerns expressed around attention and crash involvement.

3.3.3. Issues with the supervising role:

A reoccurring theme expressed by 18% (n=14/76) of the participants was that they found the supervising role to be problematic to carry out or hard to grasp. Furthermore, some participants also said that they became passive (9%, n=7/76) or that they became more of a passenger in the driver's seat (7%, n=5/76). This is exemplified by the following quotes:

"It was exciting but I don't know for how long you are able to handle it. The better it becomes, the harder it gets… This is a very big problem because when something really happens it is catastrophic when you are too far away to react. Then it is basically over."

"... somehow you cannot relax completely as you can in the passenger seat when you know that someone else has the responsibility, now it feels like a middle stage… it is a bit unclear what you should do when there is something on the road"

"I had great trust in how the car acted and got lost in my active role and towards the end I sat and wondered why I should sit here and have such a passive role [and still need to have your hands on the wheel] yes"

"When you have to sit and supervise the whole time it detracts the whole thing I would say. Because if you still need to sit and supervise and check everything then you could just as well do it yourself. Because then you have something to do in the meantime."

"It felt a bit unaccustomed in the beginning, but after a lap you got used to the thought of actually being a passenger but behind the wheel."

There was however no clear difference between crashers and non-crashers regarding reported difficulties with the supervising role.

3.3.4. Education:

The participants in experiment 3 were asked to rate to which extent the driver education prepared them for the drive ranging between 1 (not at all) and 7 (completely). The average rating was 5.63 (SD=1.45, N=16) for crashers and 6.23 (SD=1.10, N=44) for non-crashers. The participants were satisfied with the education and there was no clear difference in rating that could potentially explain crash outcome.

3.4. Driver actions and experiences during the drift event (E2)

In experiment 2, 38% (n=6/16) of the drivers did not intervene when the car drifted over into the adjacent lane. Four of these six participants did also crash with the garbage bag in the conflict event. One participant stated that she did not trust the car in the conflict event because of the earlier drift event. All participants stated post drive that they noticed the drift event during the drive.

The most common themes for non-intervening participants (note that some had statements in several themes, so those below are not mutually exclusive) were that they thought that it was a part of the test (3/6), that they would have acted sooner on a public road (2/6), that they looked ahead for obstacles instead of intervening (2/6), that they observed and got ready to intervene (2/6) and in one case the driver was afraid of aborting the automation and thus the test.

3.5. Driver actions and expectations during the conflict situation

This section presents subjective data acquired post-drive specifically related to the conflict situation.

3.5.1. Driver actions:

The subjective data results regarding driver actions are presented in table 2.

Table 2 *Themes – Driver actions E2+E3*

Theme	% (n)	% Crashers (n)	% Non-crashers (n)
I intervened	37% (28)	0% (0)	51% (28)
I intervened late	41% (31)	19% (4)	49% (27)
I intervened too late	7% (5)	24% (5)	0% (0)
I did not intervene	16% (12)	57% (12)	0% (0)

Table 2 shows that a majority of the drivers intervened. All of the non-crashers intervened or intervened late, while a majority of the crashers did not intervene.

3.5.2. Realization of the need to act:

The subjective data results regarding realization of the need to act are presented in table 3.

Table 3 *Themes – Realization of the need to act E2+E3*

Theme	% (n)	% Crashers (n)	% Non-crashers (n)
I realized the need	38% (29)	10% (2)	49% (27)
I realized the need late	30% (23)	14% (3)	36% (20)
I realized the need too late	13% (10)	48% (10)	0% (0)
I did not realize the need	18% (14)	29% (6)	15% (8)

As shown in table 3, a majority of the drivers realized the need to act. Some non-crashers expressed that they did not realize the need to act but instead acted instinctively. A majority of the crashers realized the need to act too late or not at all. Explanations to the five other crashers are presented in section 3.5.6.

3.5.3. Conflict object perception:

The subjective data results regarding perception of the conflict object are presented in table 4.

Table 4 *Themes – Perception of the object E2+E3*

Theme	% (n)	% Crashers (n)	% Non-crashers (n)
I perceived the object early	68% (52)	43% (9)	78% (43)
I perceived the object early, uncertain if in lane	8% (6)	10% (2)	7% (4)
I perceived the object late	24% (18)	48% (10)	15% (8)

Table 4 shows that all drivers perceived the object at some point in time. A majority perceived the object early, of which a few were uncertain about the lateral position of the object. Among the crashers, 48% (n=10/21) did not notice the object until it was fully revealed after the LV cut-out manoeuvre, compared to 15% (n=8/55) of the non-crashers. All of the 10 crashers that did not notice the object early did however have their eyes on-path towards the object when it was visible in the curve/crest prior to the lead vehicle cut-out manoeuvre [21].

3.5.4. Expectations on the automation:

The subjective data results regarding expectations on the automation being able to intervene in the conflict situation are presented in table 5.

Table 5 *Themes – Expectations on the automation E2+E3*

Theme	% (n)	% Crashers (n)	% Non-crashers (n)
I expected an intervention	54% (41)	95% (20)	38% (21)
I was uncertain about an intervention	28% (21)	5% (1)	36% (20)
I did not expect an intervention	18% (14)	0% (0)	25% (14)

As shown in table 5, a majority of the drivers expected the automation to intervene in the situation. All crashers except one expected the automation to intervene. Only 18% (n=14/76) of the drivers did not expect the automation to intervene.

Participants were also asked if they felt that they received enough information from the vehicle during the drive. Half of the participants (n=38/76) expressed that they wanted the car to warn them of the target and/or that they needed to intervene manually.

3.5.5. Trust in the automation

The subjective data results regarding trust in the automation vehicle being able to avoid a crash are presented in table 6.

Table 6 *Categories – Extent of trust in the automated vehicle E2+E3*

Category	% (n)	% Crashers (n)	% Non-crashers (n)
I had high trust	61% (46)	100% (21)	45% (25)
I was neutral	5% (4)	0% (0)	7% (4)
I had low trust	34% (26)	0% (0)	47% (26)

Table 6 shows that the non-crashers were close to equally represented in the high and low trust category, while all the crashers were categorized as high trusters.

In figure 4 the frequency of trust ratings separated by crash outcome are shown.

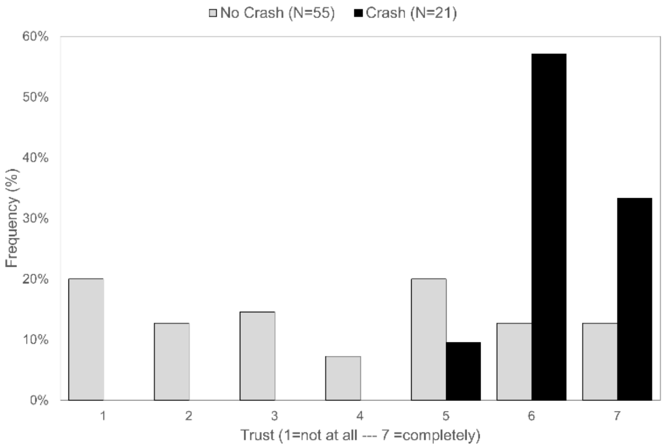

Figure 4 - *Rated trust in the automation being able to handle the situation for crashers (black bars) and non-crashers (grey bars).*

The drivers rated trust in the automation being able to handle the conflict on a medium or neutral level on average (M = 4.50, SD = 2.10, N=76). The crashers reported higher trust in the automation (M = 6.24, SD = 0.62, N=21) compared to the non-crashers (M = 3.84, SD = 2.09, N=55); t(71) = 7.68, p=0.000. There were no clear differences in trust ratings dependent on experiment, object type, or hands on wheel condition.

3.5.6. Visualization of actions and mental process

To sum up the driver actions and expectations in the conflict situation, the following figures visualize the different themes and the links between them. The charts are read from right to left, starting with the outcome which is then linked to different themes. The numbers in the boxes and on the links show the frequency of occurrence in the aggregated chart. The most common themes are marked in grey which together with the black links show the most common patterns.

Figure 5 visualizes role model performance (the “optimal process”) of a supervising driver, found in the data of 11 participants (all non-crashers).

Figure 6 visualizes the process of all 21 crashers. There are five special cases that need to be explained. Two of the four crashers that were categorized as *I intervened late* did not notice that there was any contact between the TV and the conflict object and therefore described their realization and action as late rather than too late. The third crasher was aware that a crash occurred but expressed that she did not apply enough steering power rather than intervening too late. The fourth crasher is a participant that expressed that she realized the need to act but were uncertain how to act since it was a test situation. The final special case is a crasher that expressed that she realized the need to act but chose to give the automated vehicle a chance to solve it since it was a test situation.

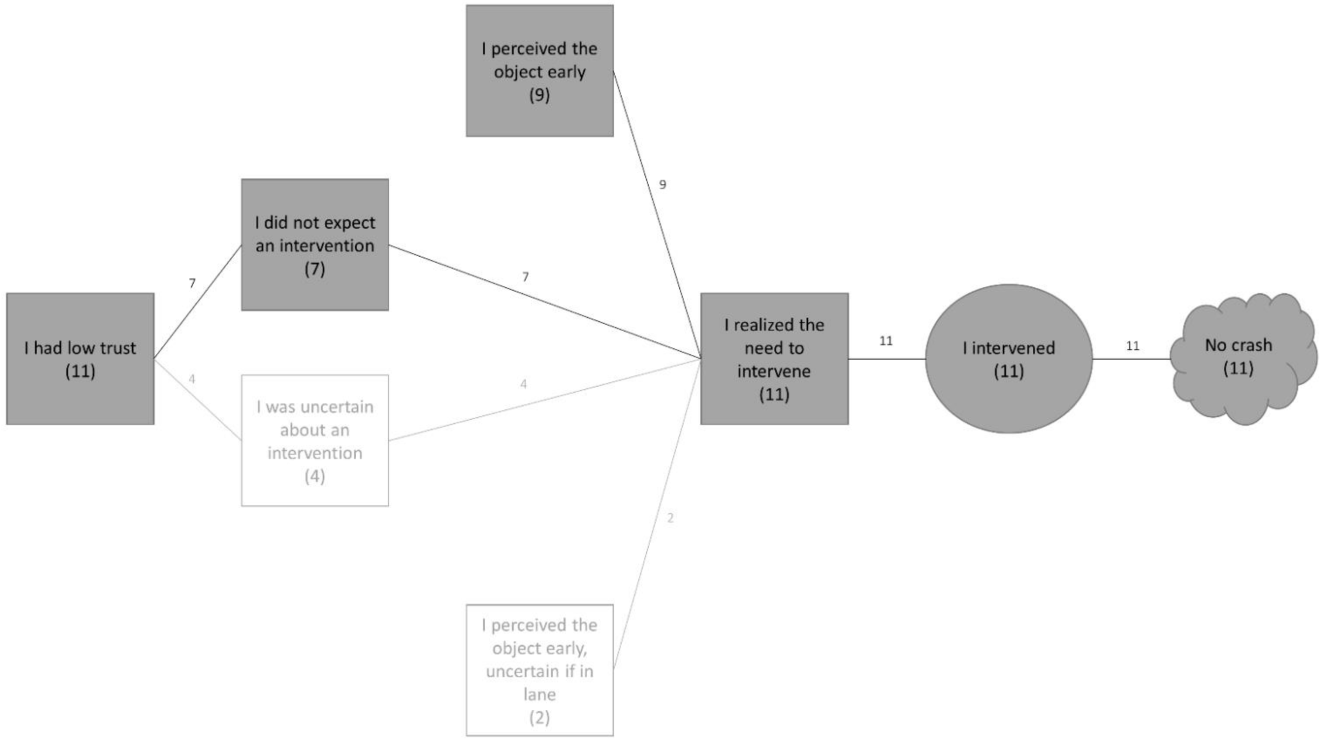

Figure 5 – *Visualization of outcome and themes for role model drivers (drivers with an optimal response in the conflict situation, n=11).*

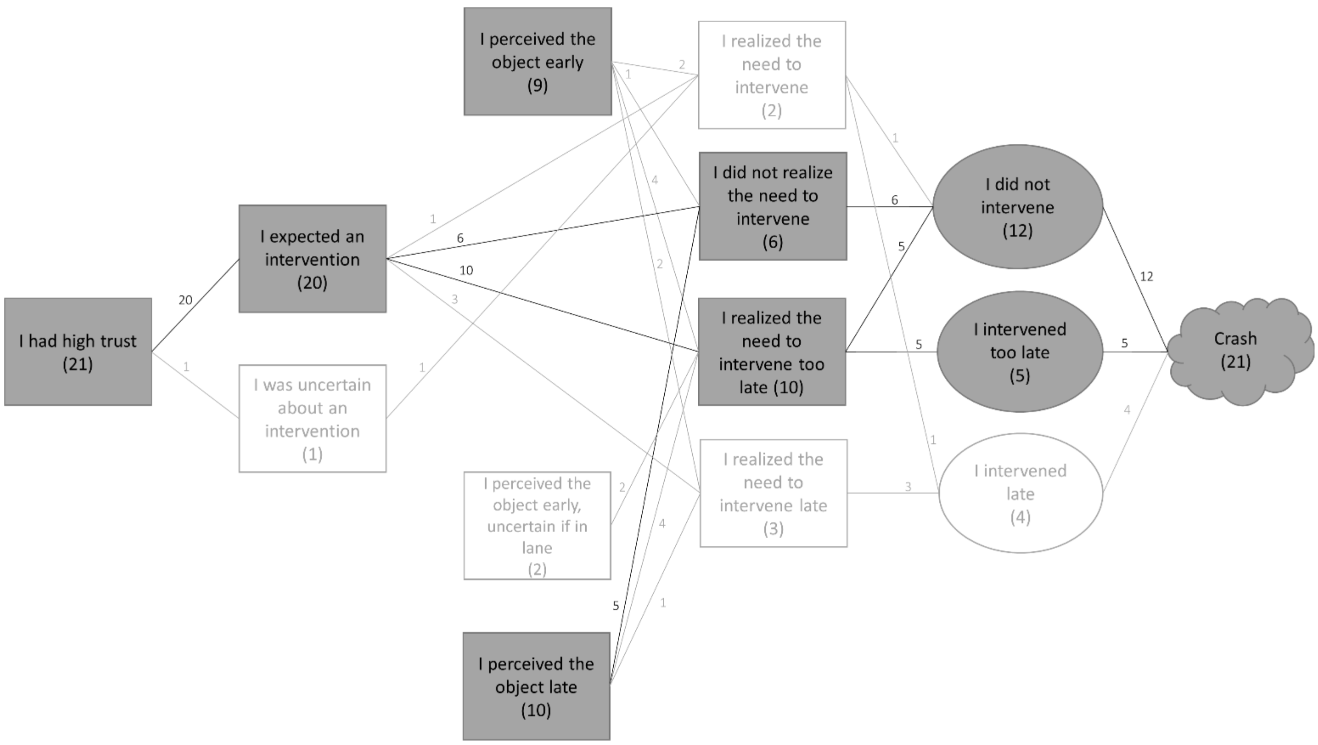

Figure 6 - *Visualization of outcome and themes for all drivers that crashed in the conflict situation (n=21).*

3.6. Overall factors affecting trust in the automation

This section presents participants' explanations to their ratings of trust in the automation being able to avoid a crash. The final themes are based on, but not limited to, the explanations of trust given when the participants rated their trust. This was supplemented by taking other parts of the interview into account, since factors affecting trust were found spread over the entire interview.

The results are presented mainly based on trust ratings independent of crash outcomes. To clarify, some participants had statements in several themes, providing themes that are not mutually exclusive. Additionally, some participants did not have any statements that were included in the final themes. To clarify the theme *Expected vehicle able to detect object and intervene*, this included explicit statements of trusting the automated vehicle based on expectations on it being able to detect the specific conflict object and/or handle the specific conflict situation as it unfolded. This theme is thus more explanatory than the more general *I expected an intervention* theme in section 3.5.4. The theme *Uncertain driver responsibility* is related to the general issues with the supervisor role found in section 3.3.3, but is strictly connected to trust and responsibility in the conflict situation (i.e. uncertainty regarding who was responsible to intervene since they believed both were able to).

The most common themes among the drivers that rated their trust as high (5-7) were that they expected the vehicle to be able to detect the object and intervene (52%, n=24/46), that they based their trust on the vehicle's good driving performance (28%, n=13/46), that they felt safe during the drive (15%, n=7/46), and that they felt uncertain about their responsibility in the conflict situation (15%, n=7/46). Note that uncertainty about responsibility in the conflict situation was expressed only by crashers; none of the non-crashers expressed the same uncertainty (see figure 7).

Comparing E2 to E3, only 13% (n=1/8) of the high trust participants in E2 based their trust on the vehicle's good driving performance, compared to 32% (n=12/38) in E3. Also, 25% (n=2/8) expressed uncertainties regarding their responsibilities in E2 compared to 13% (n=5/38) in E3.

The four drivers that rated their trust as neutral (4) did not elaborate their rating further. Looking into expectations however, three out of four were uncertain about the car intervening in the conflict, while one did not expect an intervention.

The most common themes among the drivers that rated their trust as low (1-3) were instead that they had insufficient knowledge about the capabilities of the automation (27%, n=7/26), that the driver is responsible to handle events like this (15%, n=4/26), that they felt that they wanted to handle the situation themselves (15%, n=4/26), and that they did not expect the vehicle to be able to detect the object (15%, n=4/26). Comparing the low trust participants in E2 and E3, 33% (n=2/6) stated that they did not have enough knowledge of the capabilities of the automation in E2 while 25% (n=5/20) stated this in E3. No participant with low trust in E2 expressed that it was the driver's responsibility to handle situations like this compared to 20% (n=4/20) in E3.

The most common theme related to high trust was expecting the car to be able to detect the object and intervene. One reoccurring explanation to this expectation was the perceived good driving performance of the automation. Thus, there is a relationship between driving performance, trust and expectations. In contrast, the themes of the low trust drivers were in line with the content of the education given to drivers in E3. In figure 7 the frequencies for each theme among drivers that rated their trust as high (separated by crash outcome) and low (all non-crashers) are presented.

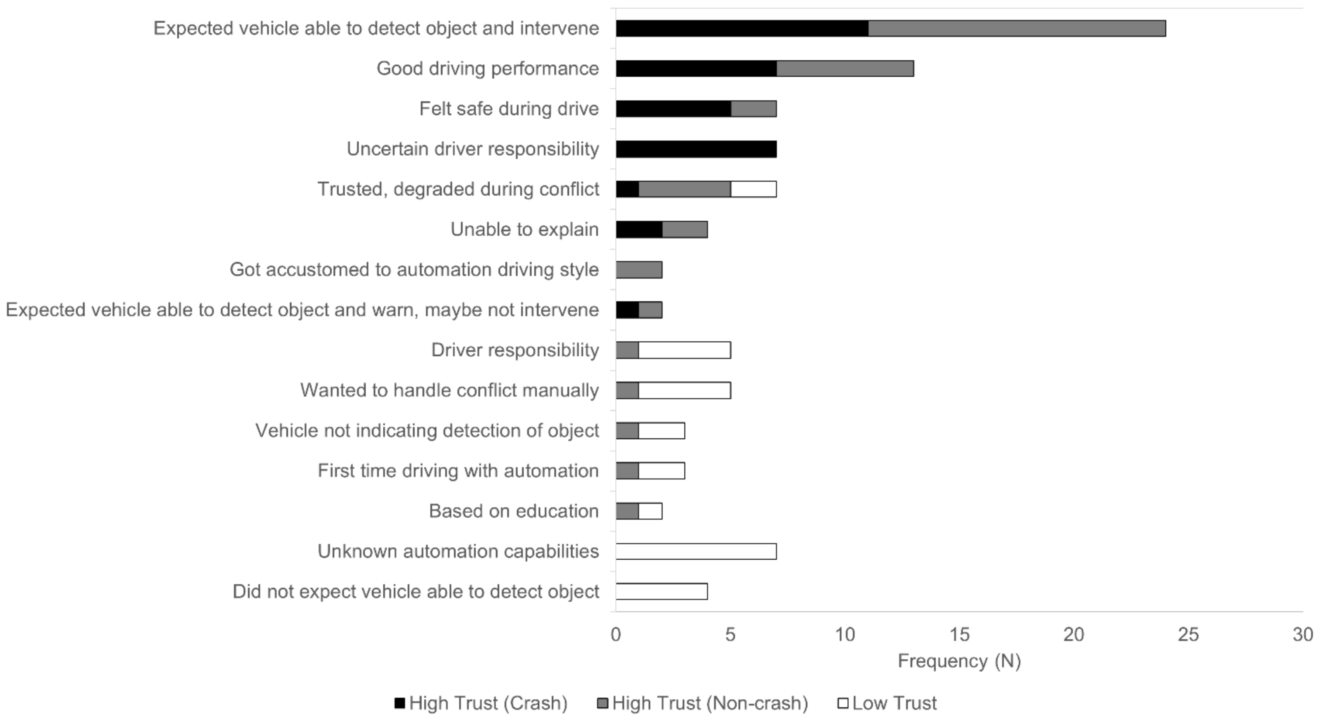

Figure 7 - *Frequency of trust themes in participants that reported high trust and crashed (black), high trust and did not crash (grey), and low trust (white).*

4. Discussion and conclusions

Explanations of the mental processes associated with the conflict event in crashers and non-crashers were highly informative. There are several interesting findings in the subjective data that warrant further discussion.

First, the behaviours and themes often brought up as concerns in the discussion of supervised automation (i.e. sleepiness, inattention, problems with the supervising role as such and pre-drive education) were not found to be highly explanatory in distinguishing between crashers and non-crashers in the subjective data from these experiments; they were quite equally expressed among participants independent of crash outcome.

Second, it certainly seems possible to engineer a supervision reminder that is both effective in terms of keeping eyes on road and hands on wheel and has high user acceptance. However, it is also evident that having eyes on road and hands on wheel does not equate to being sufficiently in the loop to act on imminent conflicts in supervised automation.

Third, interesting patterns regarding trust and expectations emerged.

To start with, a majority of the participants did realize the need to act and managed to avoid a crash. Most of these drivers reported seeing the conflict object early on, did not have very high levels of trust in the automation and were not necessarily expecting the vehicle to act. Taken together, these subjective reports correspond well with their conflict response, i.e. to actively intervene to avoid the conflict object. Interestingly though, there were only 11 role model drivers (14%) who acted and explained their actions in full accordance with the content of the instructions.

The crashers on the other hand generally reported high levels of trust in the automation, detected the conflict object later than non-crashers and expected the vehicle to deal with the conflict object on its own. In fact, almost half of the crashers did not perceive the object until it was fully revealed despite having their gaze in the direction of the object when it was visible earlier on. There was a clear expectation mismatch as crashers expected the automation to detect the object and intervene (see table 5 and figures 6 and 7). This is a subjective data pattern that intuitively correlates well with crash outcome. Also, exposing drivers to a drift event in experiment 2 did not reduce reported trust in the conflict event 15 minutes later, similar to the findings in [16].

However, there is also a group of non-crashers who do not conform to this pattern. These drivers reported high levels of trust in the automation and were expecting the vehicle to intervene, just like the crashers did. The major difference between these high trusting non-crashers and crashers is that they did not crash. Perhaps they thereby could be considered as near-crashers, but near-crashes are typically defined from physical rather than mental closeness to crash. Further investigation of this is required.

The obvious question to ask here is of course why did this sub-group not crash? What is it that distinguishes this sub-group of non-crashers from the crashers, in spite of their very similar subjective data?

The only finding in the present analysis that could provide some answers is found in the trust explanations in section 3.6, namely the uncertainty about driver responsibility in the conflict situation that a third of the crashers expressed. Since this uncertainty was not present in the data of the non-crashers, it seems like the mental models on task allocation could differ between these high trusting crashers and non-crashers.

Another possibility is that the crash outcome represents an artificial dichotomy imposed on an underlying response time continuum. Simply put, while all these drivers may have realized the need to intervene at some point in time, these who did not crash were the faster ones to do so. It is possible that combining the subjective reports with other recorded driver and vehicle data may shed further light on this issue.

Yet another possibility is that the difference stems from some deeper, underlying trait not captured in the subjective data analysed here. It is possible that the predictive processing framework [27] can provide answers and this should be further investigated. For example, the most common theme among drivers who expressed a high level of trust in the automation was expecting the vehicle to be able to detect the conflict object and intervene, despite prior training on system limitations. The most common reason for this expectation was the perceived good driving performance of the vehicle during the 30 minutes prior to the conflict. Another possible interpretation here is therefore that crashers represent drivers who are more susceptible to dynamic learned trust [17]. For them, those 30 minutes of uneventful driving in a highly reliable automated vehicle with good driving performance was enough to generate first failure effects, i.e. incorrect predictions of avoidance capability. This naturally leads to the question if more of the high trusting non-crashers would have crashed if the drive had been longer in time, or repeated on another day. That is a very interesting topic for future research.

In sum, expectation mismatch is clearly evident in subjective explanations of mental processes in drivers who crash during supervised automated driving.

5. Acknowledgments

This research was supported by the Swedish FFI project ADEST, grant number 2014-06012, within the Drive Me programme. This research complied with the tenets of the Declaration of Helsinki and was approved by the Regionala Etikprövningsnämnden i Göteborg, Dnr:369-16.

6. References

[1] B. Seppelt and T. Victor, “Potential solutions to Human Factors Challenges in Road Vehicle Automation,” *Road Vehicle Automation,* vol. 3, pp. 131-148, 2016.

[2] T. Victor, M. Rothoff, E. Coelingh,et al., “When Autonomous Vehicles Are Introduced on a Larger Scale in the Road Transport System: The Drive Me Project,” in *Automated Driving: Safer and More Efficient Future Driving*, D. Watzenig and M. Horn, Eds., Cham, Springer International Publishing, 2017, pp. 541-546.

[3] M. Lindman, I. Isaksson-Hellman and J. Strandroth, “Basic numbers needed to understand the traffic safety effect of Automated Cars,” in *IRC-17-40 IRCOBI Conference 2017*, 2017.

[4] L. Onnasch, C. D. Wickens, H. Li, et al., "Human Performance Consequences of Stages and Levels of Automation An Integrated Meta-Analysis," *Human Factors: The Journal of the Human Factors and Ergonomics Society, 0018720813501549,* 2013.

[5] R. Parasuraman, T. Sheridan and C. Wickens, "Situation Awareness, Mental Workload, and Trust in Automation, Empirically Supported Cognitive Engineering Constructs," *Journal of Cognitive Engineering and Decision Making,* vol. 2, no. 2, pp. 140-160, 2008.

[6] R. Parasuraman and D. Manzey, "Complacency and bias in human use of automation: An attentional integration," *Human Factors: The Journal of the Human Factors and Ergonomics Society,* vol. 52, no. 3, pp. 381-410, 2010.

[7] N. Strand, J. Nilsson, I. Karlsson, et al., "Semi-automated versus highly automated driving in critical situations caused by automation failures," *Transportation Research Part F: Traffic Psychology and Behaviour,* vol. 27, pp. 218-228, 2014.

[8] SAE, "SAE J3016 Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicle.," 2016.

[9] N. Merat, B. Seppelt, T. Louw, et al., "The "Out-of-the-Loop" Concept in Automated Driving: Proposed definition, measures and implications.," *Cognition, Technology and Work.,* (in prep).

[10] E. Miller and J. Cohen, "An Integrative Theory of Prefrontal Cortex Function," *Annual Reviews of Neuroscience,* vol. 24, pp. 167-202, 2001.

[11] M. Corbetta and G. Shulman, "Control of goal-directed and stimulus-driven attention in the brain," *Nature Reviews Neuroscience,* vol. 3, pp. 201-215, 2002.

[12] L. Bainbridge, "Ironies of automation," *Automatica,* vol. 19, pp. 775-779, 1983.

[13] C. Wickens, B. Hooey, B. Gore, et al., "Identifying Black Swans in NextGen: Predicting Human Performance in Off-Nominal Conditions," *Human Factors,* vol. 51, pp. 638-651, 2009.

[14] R. Molloy and R. Parasuraman, "Monitoring an automated system for a single failure: Vigilance and task complexity effects," *Human Factors,* vol. 38, no. 2, pp. 311-322, 1996.

[15] U. Metzger and R. Parasuraman, "Automation in future air traffic management: Effects of decision aid reliability on controller performance and mental workload," *Human Factors,* vol. 47, pp. 35-49, 2005.

[16] S. Hergeth, "Automation Trust in Conditional Automated Driving Systems: Approaches to Operationalization and Design," 10.13140/RG.2.2.35823.02724. , 2016.

[17] K. Hoff and M. Bashir, "Trust in automation: Integrating empirical evidence on factors that influence trust," *Human Factors,* vol. 57, pp. 407-434, 2015.

[18] D. Kaber and M. Endsley, "The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task," *Theoretical Issues in Ergonomics Science,* vol. 5, no. 2, pp. 113-153, 2004.

[19] E. de Visser and R. Parasuraman, "Adaptive Aiding of Human-Robot Teaming," *Journal of Cognitive Engineering and Decision Making: Effects of Imperfect Automation on Performance, Trust, and Workload,* vol. 5, no. 2, pp. 209-231, 2011.

[20] Y. Seong and A. Bisantz, "The impact of cognitive feedback on judgment performance and trust with decision aids," *International Journal of Industrial Ergonomics,* vol. 38, no. 7-8, pp. 608-625, 2008.

[21] T. W. Victor, E. Tivesten, P. Gustavsson, et al., "Automation Expectation Mismatch: Incorrect Prediction Despite Eyes on Threat and Hands on Wheel," *Human Factors,* 2018, (accepted).

[22] T. W. Victor, Keeping Eye and Mind on the road. Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences 9, Uppsala: Acta Universitatis Upsaliensis, 2005.

[23] J. Lee, J. Moeckli, T. Brown, et al., "Distraction detection and mitigation through driver feedback," National Highway Traffic Safety Administration, Report DOT HS 811 547A, 2013.

[24] K. Kaida, M. Takahashi, T. Åkerstedt, et al., "Validation of the Karolinska sleepiness scale against performance and EEG variables," *Clinical Neurophysiology,* vol. 117, no. 7, pp. 1574-1581, 2006.

[25] S. Elo and H. Kyngäs, "The qualitative content analysis process," *Journal of Advamced Nursing,* vol. 62, no. 1, pp. 107-115, 2008.

[26] M. Ljung Aust, A. Habibovic, E. Tivesten, et al., "Manual for DREAM version 3.2, Driving Reliability and Error Analysis Method," Chalmers University of Technology, Gothenburg, Sweden, 2012.

[27] J. Engström, J. Bärgman, D. Nilsson, et al., "Great expectations: a predictive processing account of automobile driving," *Theoretical Issues in Ergonomics Science,* vol. 19, no. 2, pp. 156-194, 2017.

Influence of non-driving-related tasks’ motivational aspects and interruption effort on driver take-over performance in conditionally automated driving

Dennis Befelein*[1], Juliane Boschet[1], and Alexandra Neukum[1]

[1] WIVW Institute for Traffic Sciences, Wuerzburg, Germany

(E-mail: befelein@wivw.de)

Abstract

Conditionally Automated Driving (CAD) as defined in SAE Level 3 (SAE, 2014) requires the driver as a fallback level in situations the car is unable to handle. The influence of non-driving-related tasks (NDRTs) on drivers’ take-over performance is an issue of ongoing debate. The study at hand analyzed subjective and objective take-over measures as a function of drivers’ task motivation achieved by the possibility to earn extra money and task interruption effort. A total of $N = 53$ participants (mean age = 32.3 years, $SD = 9.7$ years) took part in a driving simulator study with eight take-over situations. Higher task interruption effort through the instruction to store the task device in a box produced significantly longer reaction times to the Request to Intervene (RtI) with latencies between 1.5 s and 1.6 s - an equivalent of 50 meters at the implemented set speed. Although in a post-hoc rating participants considered performing the study task for incentive more critical than without external rewards, no differences between motivation conditions showed up in RtI reaction times. Results demonstrated a large impact of task interruption effort on drivers’ reaction times in SAE Level 3 take-over scenarios. High task interruption effort is a typical characteristic of real-life NDRTs that requires increased attention in future research on automated driving.

Keywords: automated driving; automation; NDRT; take-over;

Introduction

Driving automation research is a field that has increasingly gained attention within the last decade. The expected benefits of automated driving functions include increased traffic safety (e.g., through the compensation of driver deficiencies and the prevention of so-called “human errors”), the saving of energy (e.g., economizing fuel through a more balanced way of driving) as well as temporal and mental resources (e.g., by releasing the driver from the driving task and allowing him/her to relax or deal with other activities). Conditionally Automated Driving (CAD) has the potential to fundamentally change driving experience as well as driving demands in the near future. It goes one step further than Partial Automation (which is already available on the market by several automobile manufacturers) by relieving the driver from the obligation to continuously monitor the driving environment and system status of the vehicle. Instead, it suffices if he/she is able to respond to a possible take-over request within an adequate period of time (Gasser et al., 2012; NHTSA, 2013; Pfleging, Rang, & Broy, 2016; SAE, 2014). At this level, take-over requests are expected to occur only when system limits are reached. Prominent examples being work zones, highway endings, missing lane markings or system failures. With the necessity of system monitoring being dropped, non-driving-related tasks (NDRTs) that had been distracting or even forbidden during partially automated driving are back on stage and require reassessment.

Comparative studies have shown that different NDRTs produce different take-over outcomes in terms of reaction times and take-over quality (Naujoks, Purucker, Wiedemann, & Marberger, submitted; Vogelpohl, Vollrath, Kühn, Hummel, & Gehlert, 2016). This raises the question if there are higher-level task characteristics that influence drivers’ availability in take-over situations. Standardized NDRTs are widely used in automation research for their easy manipulation, reproducibility and adequacy to measure psychological constructs like cognitive workload or distraction. Studies using standardized NDRTs (like e.g., the Surrogate Reference Task or the n-Back-task) provided evidence that driver take-over behavior is influenced by modality of the NDRT (Gold, Berisha, & Bengler, 2015) and traffic situation (Radlmayr, Gold, Lorenz, Farid, & Bengler, 2014). The study at hand uses a different approach by using more naturalistic NDRTs that are closer to realistic driving situations and differ in the rather practical aspects of task motivation and interruption effort.

Public opinion studies indicate that future users of automated driving will engage in motivating tasks, such as texting, eating/drinking, surfing the internet or watching movies (Pfleging et al., 2016; Schoettle & Sivak, 2014). Evidence from real traffic research points into the same direction, showing that drivers who have experience with driving assistance functions show increased secondary task engagement during partially automated driving (Naujoks, Purucker, & Neukum, 2016). Since it is widely accepted that motivating tasks are preferably continued than monotonous ones, we assume that drivers with enhanced task motivation will show longer take-over reaction times and poorer take-over quality than those with lower motivation to continue the task.

Naturalistic NDRTs may also differ from standardized ones in terms of interruption effort, which refers to necessary motoric steps to interrupt the respective NDRT and lay related objects aside. Complex physical tasks like e.g., eating or reading a large newspaper may be hard to interrupt since related objects may have to be cleared away with effort. We therefore suppose that drivers who are engaged in tasks with high interruption effort will show longer take-over times and poorer take-over quality than those engaged in a task with low interruption effort.

The present study compares two different motivational driver states regarding the NDRTs, as well as two differently effortful interruption conditions of these tasks. The impact of these manipulations on drivers' take-over behavior will be investigated at the example of a broken-down vehicle on the ego-lane.

Method

Driving simulation

The study was conducted in the static high-end driving simulator (Figure 1) of WIVW GmbH. The driving simulation software SILAB was used for environment visualization as well as for simulation of the ADAS for cooperative driving, traffic and vehicle dynamics. An Opel Insignia Sports Tourer is used as mockup of the driving simulator. The simulator had a 300° horizontal and 47° vertical field of vision, with five image channels, each one with a resolution of 1400x1050 pixels. The update frequency was 60 Hz. In addition, there were two LCD displays representing the right and left outside mirror. The interior mirror reflects a LCD display positioned in the trunk of the mockup showing the scenery behind the vehicle. During the experimental drives, the experimenter was able to observe the driver and to communicate with the participant via intercom.

Figure 1. The static WIVW driving simulator.

Conditional automation specifics

Vehicle automation included lateral and longitudinal guidance (SAE Level 3) with a set speed of 120 km/h. Set speed was reached whenever there was no slower vehicle ahead. Within automated driving sections, no lane changes were executed, neither by the system nor by the driver. In case there were slower vehicles ahead, they were followed with a pre-set time-headway of 2 s. The system was activated and deactivated via pressing two steering wheel buttons simultaneously that could easily be reached with the driver's thumbs when holding the wheel at "ten and two". Steering against the counterforce of the automation at a steering wheel angle larger than 2° also deactivated the automation. Lane changes were not necessary during CAD sections.

Scenario Layout

In the study at hand, an emergency take-over request on a highway was examined. The ego-vehicle was driving autonomously on the right lane following a lead vehicle at 120 km/h. At a predefined point, the lead vehicle pulled out to the left and gave view to a broken-down vehicle on the ego-lane. At the same moment, a visual-auditory Request to Intervene (RtI) was issued and longitudinal guidance was shut off, leading to drag torque related deceleration. Time to collision (TTC) at the moment of RtI output was approx. 9 s. An absence of driver reaction would have resulted in a collision with the standing car.

Human Machine Interface

The RtI was visualized in the vehicle's central infotainment display (Figure 2). It disappeared when the driver deactivated the system by braking or pressing the buttons (as described above). The visual display was accompanied by two consecutive high frequency warning tones to prompt immediate driver intervention.

Study Design

A complete within-design was used in the study. Every participant completed two blocks in randomized order: A block with the NDRT for external incentive and a block with the NDRT as a simple pastime. Both blocks further split up into two consecutive take-over situations with high and two consecutive take-over situations with low interruption effort. As a result, every participant encountered eight takeover situations.

Figure 2. The visual RtI from the vehicle's central infotainment display.

Independent Variables

The video game Tetris® was chosen as NDRT because it could hardly be neglected by the driver without score loss, thereby requiring continuous task attention. However, the game could be paused with a "Pause"-button on the tablet screen. The game was provided on two identical eight inch hand held Samsung tablets which for better discriminability of the motivation conditions were color coded. Driver motivation was manipulated by external rewards: When playing with the yellow tablet, drivers could monitor their high score and were instructed to give their best to earn extra money depending on their performance (performance condition). For every Euro earned, a cash register sound was presented, and the actual profit was reported verbally by the experimenter via intercom. When playing with the red tablet, drivers could neither see their high score, nor could they win any money, and the experimenter described the task as a simple pastime without any performance measurement (pastime condition).

Task interruption effort was manipulated by two different interruption instructions: To create high interruption effort, drivers were instructed to pause their task on the tablet, put the device into a plastic box on the co-driver's seat and place a lid on top of the box before taking over vehicle control (Figure 3). For low task interruption effort, it sufficed to pause the tablet task and lay the device aside, but not into the box. Continuous task processing and correct interruption were monitored by the experimenter.

Figure 3. Box for high task interruption effort (with lid and tablet).

Dependent Variables

On an objective level, the time from RtI onset to the first driver reaction was of particular interest. It was defined as the first of the following driver reactions: (1) System deactivation with the steering wheel buttons, (2) braking, or (3) steering with more than 2° steering wheel angle.

On a subjective level, drivers were asked to rate the criticality of the take-over situations directly after they had completed them using the 'scale of criticality assessment of driving and traffic situations' (Figure 4). The scale was originally developed in order to assess the controllability of erroneous interventions of driver assistance systems (Neukum & Krüger, 2003) and later extended to the assessment of the criticality of driving situations (Neukum, Lübbeke, Krüger, Mayser, & Steinle, 2008). The advantage of the scale is the definition of a threshold value that defines critical situations from the driver's perspective (rating as 'dangerous' or 'uncontrollable').

Figure 4. Scale of criticality assessment of driving and traffic situations.

In addition, directly after each take-over, drivers rated helpfulness of the TOR as well as their own take-over performance on Likert scales ranging from 0 to 15. At the end of the study, subjects filled out a questionnaire related to task involvement which contained similar Likert scales. The questionnaire also served as a manipulation check. Items were:

- *"How pronounced was your motivation to play Tetris?"* (subsequently referred to as 'task motivation')
- *"How hard was it for you to interrupt the game*?" (referred to as 'hardness to interrupt')
- *"How critical do you consider playing Tetris during a real, highly automated freeway drive?"* (referred to as 'task criticality')

Items had to be answered separately for conditions with and without monetary reward.

Procedure

Upon arrival, subjects were welcomed and gave informed consent. The experimenter explained that the goal of the study was the evaluation of a visual display under different distraction conditions. In a next step, the functionality of the conditional automation was explained. Participants were instructed that they did not have to monitor driving when the automated system was active and should fully apply themselves to the NDRTs. They were told that whenever they had to take back vehicle control, they system would inform them in time. The different motivation and interruption conditions were explained as well. The training was rounded off with a short drive in which participants practiced system (de)activation and the two interruption conditions without encountering take-over requests. It was finished when participants had fully understood system operation as well as motivation and interruption procedures.

The following main drive consisted of eight highly automated driving sections that each lasted approx. 3 min and were followed by the previously explained take-over situations. The test course was designed in a way that take-over situations were hardly predictable for the drivers. The experimenter instructed which tablet was to be used before the respective takeover situations, and how the task had to be interrupted in case of a possible take-over request. When subjects started a "performance task" section, they were also verbally motivated by the experimenter (*"Now try to give your best and become high score leader!"* etc.). When they started a "pastime task" section, verbal instructions were kept explicitly discouraging (*"Now you can start playing as a pastime, but your performance doesn't matter."* etc.). After the main drive, participants completed questionnaires, received monetary compensation for their participation, and were discharged. The entire procedure took approx. 40 min.

Participants

A total of N = 58 participants took part in the study. 28 participants were female and 30 male. The mean age was 32.3 years (SD = 9.7 years). The oldest driver was 54 and the youngest driver 19 years old. Participants were recruited from the WIVW test driver panel and had taken part in an extensive driving simulator training (Buld, Krüger, Hoffmann, & Totzke, 2003) prior to the study.

Data exclusion

Driving data results revealed training effects between the first two take-over situations across participants, so the first of the eight take-over situation of every subject was excluded from the analysis. Of the 406 take-over situations analyzed, 36 had to be reclassified because participants confused the instructed interruption conditions. For example, when participants in a condition with “high interruption effort” only laid the tablet on the seat although they were instructed to put it into the box before taking over, the situation was reclassified into “low interruption effort”. In addition, 14 take-over situation had to be excluded because participants did not play Tetris at the moment of take-over (e.g., because they had gone game over right before).

Statistical procedure

Statistical tests were conducted using IBM SPSS Statistics Version 25. The obtained data was analyzed descriptively before applying inferential statistics. Comparisons between the manipulation conditions were realized using univariate analyses of variance. Although the present study had a *within* design, repeated-measures analyses of variance would not have been an adequate procedure because of missing data (see above). For that reason, two-factorial univariate analyses of variance without repeated measures were calculated. These analyses can be considered conservative since they do not take individual differences between participants into account.

Results

Subjective Data

Figure 5 shows mean situation criticality ratings of the take-over situations. In take-over situations with high interruption effort criticality ratings were significantly higher (M = 5.08, SD = 2.49) than in situations with low interruption effort (M = 3.75, SD = 2.12; F (3,404) =

17.30, $p < .001$). There were no significant differences between motivation conditions nor any interactions.

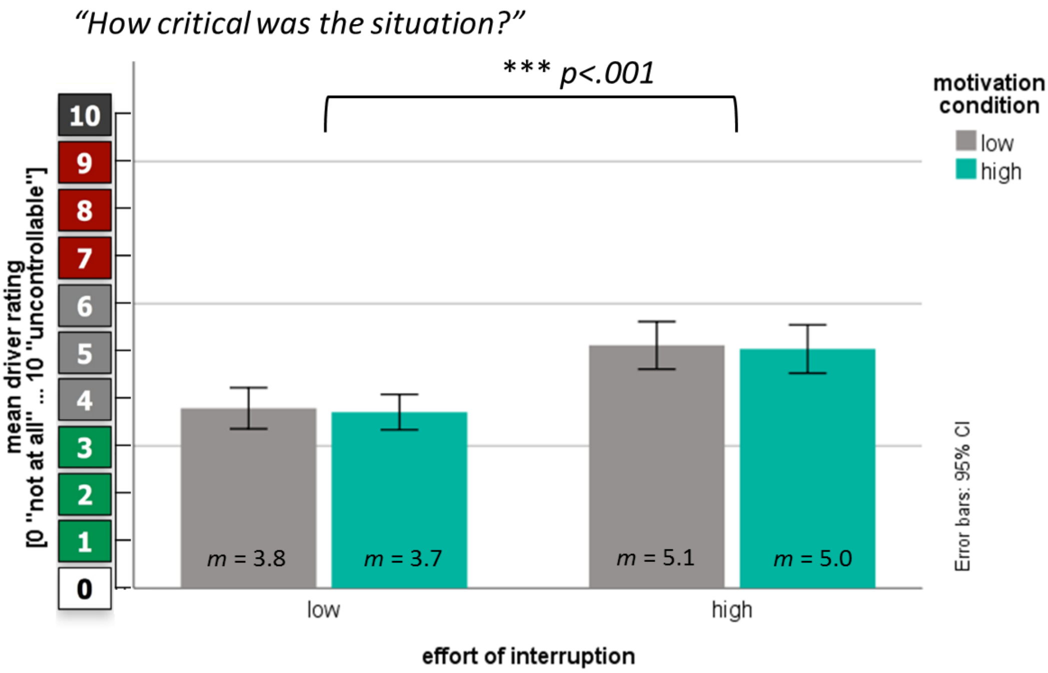

Figure 5. Mean criticality ratings gathered directly after the take-over situation, as a function of motivation condition and effort of interrruption.

The interruption effort also influenced drivers' self-rated take-over performance (Figure 6). Although all ratings were in the range from 10 to 12 (*"good"*), drivers rated their take-over performance significantly lower in situations with high interruption effort ($M = 10.6$, $SD = 2.8$) than in situations with low interruption effort ($M = 11.9$, $SD = 2.0$; $F(1,406) = 31.59$, $p < .001$). There were no significant differences between motivation conditions nor any interactions.

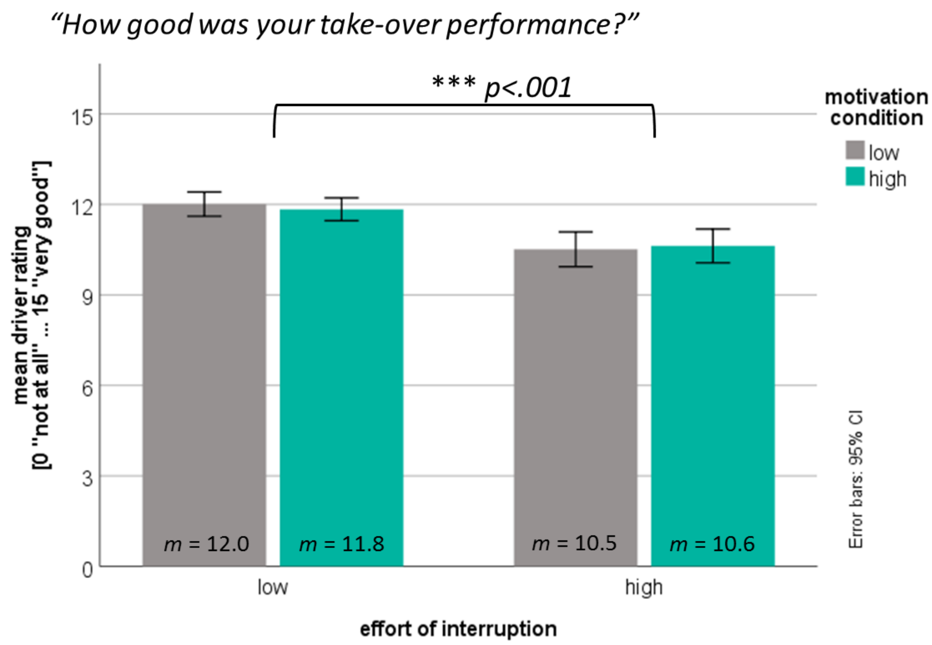

Figure 6. Mean self-reported driver performance ratings gathered directly after the take-over situation, as a function of motivation condition and effort of interruption.

In an inquiry after the test drive, drivers had to give their degree of agreement to the statement "How dangerous do you consider playing Tetris during real, highly automated highway drives?" on a 15-point Likert scale. The performance condition (with high score and money) was considered significantly more dangerous (M = 11.2, SD = 3.0) than the pastime condition (M = 9.8, SD = 3.3; $F(1,128) = 6.13$, p = .015). When drivers had to rate their task motivation and how hard it was to interrupt the playing, only minor differences between conditions occurred on a descriptive level (Figure 7).

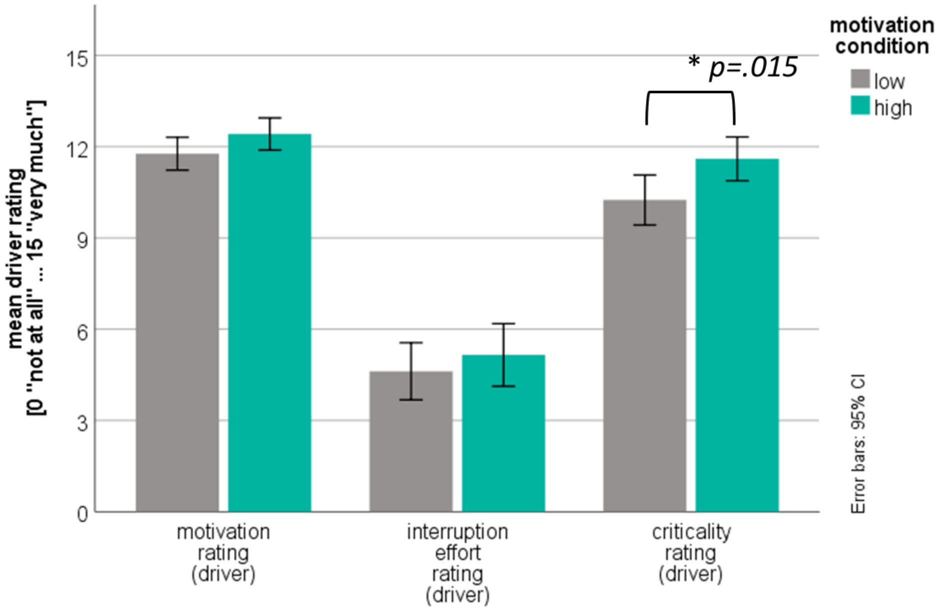

Figure 7. Subjective driver rating of task motivation, hardness to interrupt and task criticality depending on motivation condition.

Objective Data

Figure 8 shows the time to first driver reaction after the RtI (defined as previously described). In situations with high interruption effort, drivers reacted significantly slower (M = 5.3 s, SD = 1.3) than in those with low interruption effort (M = 6.9 s, SD = 1.1; $F(1,383) = 158.93$, $p < .001$). For situations with low manipulated driver motivation, mean reaction times were 5.3 s in the low interruption effort condition (SD = 1.4) and 6.8 s in the high interruption effort condition (SD = 1.0). For situations with high manipulated driver motivation, mean reaction times were 5.4 s in the low interruption effort condition (SD = 1.3) and 7.0 s in the high interruption effort condition (SD = 1.2). There were no significant differences between motivation conditions nor any interactions. The most prominent first driver reaction was button press (47.1% of all take-over situations), followed by braking (46.9%) and steering (6.0%), with very little variation within participants.

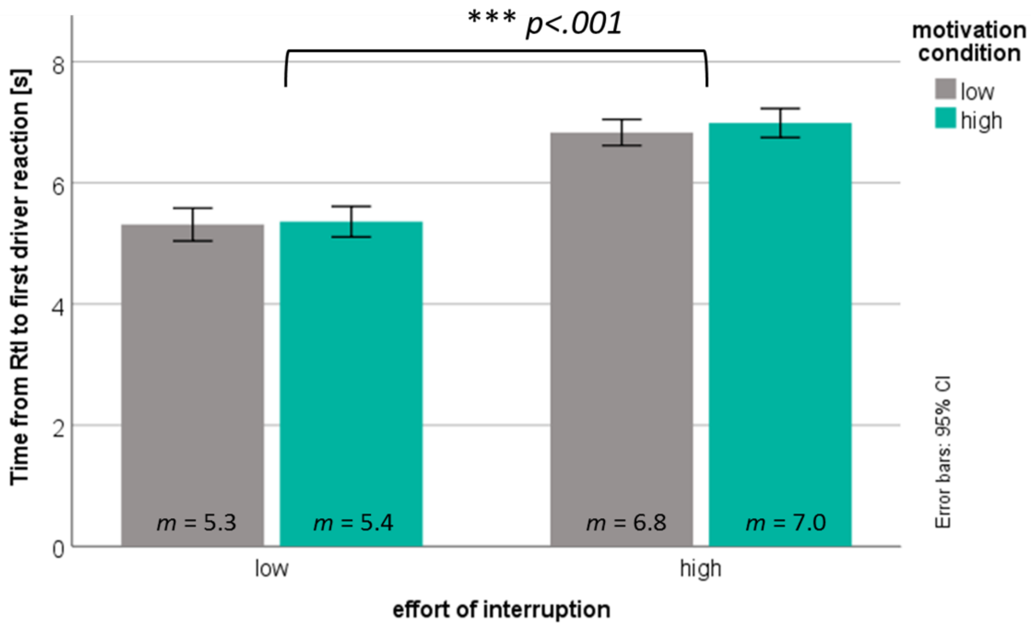

Figure 8. Mean driver reaction times following take-over requests, depending on motivation condition and interruption effort condition.

Discussion

With a large body of research focusing on easily interruptible standardized NDRTs, motoric interruption steps which are rather typical for naturalistic NDRTs have largely been disregarded. The study at hand analyzed subjective and objective take-over measures as a function of driver task motivation and task interruption effort. It could be demonstrated that task interruption effort has a considerable influence on driver take-over reaction times. Storing the task device in a box came along with significantly longer reaction times to the RtI in a range between 1.5 s and 1.6 s, an equivalent of roughly 50 meters at the implemented set speed of 120 km/h.

Considering the finding that drivers of conditionally automated vehicles are likely to engage in complex natural tasks (Pfleging et al., 2016), task interruption effort requires increased attention in future research on automated driving. Different approaches could be taken to address the issue: For example, tasks with excessive interruption effort may in part be *prevented* by limiting media use to in-vehicle screens and touch pads which do not have to be cleared away as it is the case with brought-in media devices. Additionally, these in-vehicle devices offer the opportunity to stop any visual presentation in case of RtIs (often called “lock-out”). Storage aids for NDRT devices may also help to reduce interruption effort. A second approach would be to *manage* interruption effort issues by detecting potentially critical tasks with eye tracking and posture detection, allowing to adjust RtI timing to the particular situation. However, in conditionally automated driving there will always be sudden time-critical take-over situations like in the study at hand that leave virtually no room for RtI timing adjustment.

Regarding task motivation, playing the tablet game for points and money was considered more critical by participants than playing without external rewards in the post-hoc rating. However, no differences between motivation conditions showed up in RtI reaction times. A possible explanation for this finding is provided by the manipulation check: Driver-reported motivation to play Tetris was high – almost independently from monetary incentives.

Literature

Buld, S., Krüger, H.-P., Hoffmann, S., & Totzke, I. (2003). Die Fahrsimulation als Methode bei der Entwicklung von Fahrerassistenzsystemen. *VDI-Berichte*, 153-175.

Gasser, T. M., Arzt, C., Ayoubi, M., Bartels, A., Bürkle, L., Eier, J., . . . Huber, W. (2012). Rechtsfolgen zunehmender Fahrzeugautomatisierung. *Berichte der Bundesanstalt für Straßenwesen. Unterreihe Fahrzeugtechnik*(83).

Gold, C., Berisha, I., & Bengler, K. (2015). *Utilization of Drivetime–Performing Non-Driving Related Tasks While Driving Highly Automated.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Naujoks, F., Purucker, C., & Neukum, A. (2016). Secondary task engagement and vehicle automation–Comparing the effects of different automation levels in an on-road experiment. *Transportation research part F: traffic psychology and behaviour, 38*, 67-82.

Naujoks, F., Purucker, C., Wiedemann, K., & Marberger, C. (submitted). Effects of different non-driving related tasks on take-over time and quality in non-critical situations – an on-road study *Human factors*.

Neukum, A., & Krüger, H.-P. (2003). Fahrerreaktionen bei Lenksystemstörungen–Untersuchungsmethodik und Bewertungskriterien. *VDI-Berichte*, *1791*, 297-318.

Neukum, A., Lübbeke, T., Krüger, H.-P., Mayser, C., & Steinle, J. (2008). *ACC-Stop&Go: Fahrerverhalten an funktionalen Systemgrenzen.* Paper presented at the 5. Workshop Fahrerassistenzsysteme-FAS.

NHTSA. (2013). Preliminary Statement of Policy Concerning Automated Vehicles. Washington, DC. *Verfügbar unter www. nhtsa. gov/staticfiles/rulemaking/pdf/Automated_Vehicles_Policy. pdf*.

Pfleging, B., Rang, M., & Broy, N. (2016). *Investigating user needs for non-driving-related activities during automated driving.* Paper presented at the Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia.

Radlmayr, J., Gold, C., Lorenz, L., Farid, M., & Bengler, K. (2014). *How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Chicago, Illinois.

SAE. (2014). Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. *SAE Standard J3016*, 01-16.

Schoettle, B., & Sivak, M. (2014). Public opinion about self-driving vehicles in China, India, Japan, the US, the UK, and Australia.

Vogelpohl, T., Vollrath, M., Kühn, M., Hummel, T., & Gehlert, T. (2016). Übergabe von hochautomatisiertem Fahren zu manueller Steuerung. *Berlin: GDV*.

By what hubris? – the readiness of the human operator to take over when the automation fails or hands over control

Jerry Wachtel[1*]

[1] The Veridian Group, Inc. 567 Panoramic Way, Berkeley, CA, USA
[*]jerry@veridiangroup.com

Abstract: As Level 2 automated vehicles become pervasive in the traffic stream and as Levels 3 and 4 vehicles become increasingly common, automation failures and sudden handoffs due to coding errors, unanticipated events, or hacking will also increase. Despite some encouraging findings we argue that a non-trivial percent of drivers will be ill-equipped to handle such situations. We demonstrate that, in three highly technological industries with better prepared operators, better controlled working environments, and more rigorously designed and tested equipment, accidents and near misses (incidents) still often occur during automation failures and handoffs, as well as due to the operators' misunderstanding of the automation or the state of the equipment. We express our opinion that specialized driver training and/or "chatty" on-board interfaces may be potential solutions to this problem, and that there is little or no evidence that either of these methods is in use or contemplated in the field. Finally, we propose a thought experiment to test our hypothesis about the viability of these two approaches.

1. Introduction

The U.S. based Society of Automotive Engineers (SAE International) has identified six levels of vehicle automation, supplanting the previous listing by the U.S National Highway Traffic Safety Administration (NHTSA) [1]:

Level 0 - Zero autonomy – the driver performs all driving tasks.

Level 1 – Driver Assistance – the vehicle is controlled by the driver, but some driving assist features may be included in the vehicle design.

Level 2 – Partial Automation – vehicle has combined automated functions, like acceleration and steering, but the driver must remain engaged with the driving task and monitor the environment at all times.

Level 3 – Conditional Automation – driver is a necessity, but is not required to monitor the environment. The driver must be ready to take control of the vehicle at all times with notice.

Level 4 – High Automation – the vehicle is capable of performing all driving functions under certain conditions. The driver may have the option to control the vehicle.

Level 5 – Full Automation – the vehicle is capable of performing all driving functions under all conditions. The driver may have the option to control the vehicle.

Despite ongoing trials in several countries, fully automated vehicles are not likely to become commonplace on our roadways for many years to come. NHTSA estimates that the "highway autopilot" with "fully automated safety features" will become widely available from 2025 [2]. Others are far more pessimistic [3, 4].

If and when full automation becomes commonplace, it is widely agreed that it will bring about substantial benefits to society: in increased fuel economy, reduced air pollution, travel efficiencies, and, most of all crash reduction and injury prevention. Prior to that time, however, Level 2 automation has become increasingly common, and Level 3 automation is beginning to be introduced in a limited but growing number of high end production vehicles. Since we have years to go before Level 5 may be achieved on a widespread basis, human factors experts and vehicle designers are concentrating their attention on Levels 3 and 4. This paper argues that we introduce vehicles with such increasing levels of automation with considerable hubris, based on results from other industries and growing experience with such vehicles “on the streets.”

By presenting exemplar accidents and incidents that have occurred with automation in other industries, and then comparing operations and operators in those industries to the automotive environment, we hope to point out why we believe that we are engaged in hubris, and then propose a thought experiment in an effort to address the major concerns that we see.

2. Automation Failures and Hand-offs in Other Industries

2.1 Aerospace – The Apollo 10 Anomaly

In April of 1969, the U.S. launched its final rehearsal space mission for the ultimate goal of landing the first man on the moon, which was to take place three months later. On this last rehearsal flight, identified as Apollo 10, the moon landing vehicle (called the Lunar Module, or LM), was to separate from the Command Module (CM) which remained in orbit some 60 miles above the lunar surface. The LM was then to descend to 10 miles above the surface, perform certain mission related objectives, and then fly back to complete a “rendezvous and docking” with the CM. By completing its activity, the LM would complete every step of the actual landing except the final descent and touchdown on the lunar surface. But upon ascent to rendezvous, while testing the Abort Guidance System (AGS) something went wrong. The mission Commander, who was flying the LM, complained that, when he put the Rendezvous Radar switch into the “Automatic” mode, the LM began to gyrate wildly. He quickly put the switch into the “Off” position to gather his wits. When he put it back into “Automatic” again, the spacecraft performed exactly the same way, and he was very confused and quite angry. It was only via instruction from flight controllers on the ground in Houston, Texas and Bethpage, New York, that he was able to put the switch into the “Attitude Hold” mode and fly the LM manually to achieve radar contact with the CM and, ultimately, to achieve a successful docking. The net result was the need to fly an additional lunar orbit, a very angry astronaut, a contentious flight debriefing, and a forced delay of the next flight, the first manned lunar landing, while the LM’s manufacturer (Grumman Aerospace Corporation) undertook a very public and painful analysis of the “failure” as ordered by NASA. This analysis was all the more painful because both NASA and Grumman knew that there was no problem – the spacecraft had performed exactly as it was supposed to; the telemetry data proved it. But in those days astronauts were considered national heroes who could do no wrong. And although engineering and human factors staff at both organizations knew that the Commander had erred, that he had misunderstood how the “Automatic” function worked, and, as a result, placed the vehicle into an unintended flight mode, no one would call him on it, and so the Grumman team spent two months investigating a non-event at its customer’s direction. In the end, the switch was fitted with a guard “to prevent inadvertent actuation.” Some folks were mollified; most were not. But the program went on, and the manned lunar landing was successfully performed during Apollo 11.

The following is from the NASA Apollo 10 Mission Report [5]:

“… lunar module attitudes deviated from expected during the staging maneuver. Telemetry data indicated the automatic mode was engaged twice for short periods prior to and at staging. Since the automatic mode had been used previously to point the lunar module's Z-axis at the command module, the guidance system returned the vehicle to that attitude. While considerable deviation in attitude was experienced temporarily, no adverse effects on the rendezvous resulted.” (p. 4-3).

In Section 15.2 of the report, anomalies related to the Lunar Module are discussed. Anomaly 15.2.14 addressed “attitude anomalies at staging.” “Large attitude excursions occurred prior to and during staging. Body rates of 19 deg/sec in pitch and greater than 25 deg/sec in roll and yaw were recorded. Smaller attitude excursions occurred approximately 40 seconds prior to staging. The mode switching, telemetry, and associated attitude commands indicated that the abort guidance mode changed from ATT HOLD to AUTO coincident with the vehicle gyrations. … it is considered highly remote that switch malfunctions could have caused the anomalies at staging. … It is … concluded that the anomaly was caused by the inadvertent cycling of the abort guidance mode control switch, followed immediately by an incorrect output of the yaw rate gyro. … the abort guidance mode control switch was transferred to the AUTO position, resulting in high vehicle rates during the staging sequence.”

2.2 Nuclear Power Industry – the Chernobyl Catastrophe

On April 25-26,1986, Unit 4 of the Chernobyl Nuclear Power Plant near Kiev, Russia, was being powered down for routine maintenance. While this process (which takes many hours) was underway, the operating crew initiated an experiment which had been previously attempted unsuccessfully. This test involved simulating a “station blackout” (loss of all offsite power), during which safety systems were intentionally switched off to test whether the plant’s turbines, while spinning down to idle speed, could provide intermediate power to the backup diesel generators which were to provide power to the plant (onsite power) during the blackout. As stated above, the test had been tried unsuccessfully at least three times in the past, but it could only be performed during a planned power outage which only occurred for maintenance or fuel replacement every several months.

Despite their robust training and preparation, and following their detailed procedures, the crew was not aware of two flaws in the design of the RBMK reactor, and this contributed directly to the accident. The first was that this reactor design was unstable at low power levels; the second was that, for the first few seconds of control rod insertion (a procedure used to stop a nuclear reaction), reactor power actually increased rather than reduced as desired. There is also evidence, as recorded by a centralized (remote) control system, that an emergency shutdown of the reactor was initiated when the “EPS-5 button was pressed – this fully inserted all control rods, some of which had been withdrawn earlier” [6]. This action was wrong and proved to be the immediate trigger for the subsequent initial explosion.

Over a period of nine hours, the reactor became unstable and the crew “lost control” of it. The reactor overheated, melting the nuclear fuel and causing a series of steam explosions that tore off and lifted the 2,000-ton metal plate over the rector, blew the roof off the building, and spewed radioactivity for hundreds of miles, causing radioactive particles to be carried by prevailing winds into Western Russia and Eastern Europe.

Two deaths were recorded in the facility, 134 first responders were hospitalized, of whom 28 died of acute radiation poisoning, and 14 more died of radiation induced cancers. In addition, 15 childhood thyroid cancer deaths were recorded. Russia immediately evacuated the nearest town of Pripyat, where most of the plant’s employees and their families lived. That city has been permanently abandoned and its occupants resettled. A concrete sarcophagus has been erected over the ruined facility. This, the worst disaster to confront the nuclear industry (until the 2011 meltdown at the Fukushima Daiichi nuclear plant in Japan caused by an earthquake and resultant tsunami), was caused by a highly trained crew failing to understand the behaviour of automated systems within the plant and failing to respond appropriately when these systems began to become unstable. That a “safety-related” switch was also erroneously pressed, immediately triggering the initial explosions which ultimately led to the reactor core meltdown, is further evidence of the workers’ misunderstanding of the consequences of their actions during takeover from an automated system.

2.3 Aviation Industry – The Crash of Asiana Airlines Flight 214

Asiana Airlines flight 214 was a transpacific flight from Incheon International Airport near Seoul, South Korea to San Francisco International Airport. It crashed during the final approach to landing on July 6, 2013. It was the first crash of a Boeing 777 aircraft involving fatalities since that aircraft was entered into service in 1995.

The flight was cleared for a visual approach to the runway at 11:21 am, and again at 11:27. The weather was fine. There was light wind, no precipitation, and no reports of wind shear. Visibility was 10 miles – the maximum that the system could report.

The aircraft crashed into the seawall short of the runway at 11:28 am. Both engines, the tail section, and the main landing gear separated from the fuselage upon impact. After skidding along the runway, the aircraft came to rest some 2,400 feet from the initial point of impact.

The three flight crew members had extensive flying experience. The pilot in command (who also served as a check/instructor captain, had over 12,000 hours of flying experience, of which over 3,000 were in a Boeing 777 aircraft. (12,000 hours at a driving speed of 62 mph (100 km/hour) would equate to driving 740,000 miles (1.2 million km) The captain receiving his training had nearly 10,000 hours of flight experience, of which 43 were in a 777 over nine flights.

The final report of the National Transportation Safety Board (NTSB) was issued on June 24, 2014 [7]. The Board determined that the probable cause(s) of the accident were: “the flight crew’s mismanagement of the airplane’s descent during the visual approach, the (pilot’s) unintended deactivation of automatic airspeed control, (and) the flight crew’s inadequate monitoring of airspeed… “ Contributing factors included: “the complexities of the autothrottle and autopilot flight director systems that were inadequately described in Boeing’s documentation and Asiana’s pilot training, which increased the likelihood of mode error; (and) the flight crew’s nonstandard communication and coordination regarding the use of the autothrottle and autopilot flight director systems.”

We have highlighted the Asiana crash because it is recent and has been in the news, and because it is a representative example of crashes (and near misses) that are the focus of this paper – the operators’ failure or inability to understand the automation to a sufficient degree to take over when the automation fails or needs to hand off control. But Asiana is just one of many recent aviation examples that represent such a condition. In a recent report, Mumaw [8] has compiled brief descriptions of 42 aviation accidents and events relating to “autoflight” use and misuse. While some of these incidents date to the 1970s, the vast majority have occurred within the last 20 years, when this technology became more prevalent. Some of the event categories bear a strong resemblance to concerns about autonomous vehicles: The autopilot (or autothrottle) is off or failed and the pilot thought it was engaged; the autopilot takes an action that the pilot is not aware of; the autopilot reverts to another mode; the pilot does not understand the mode’s behaviour.

2.4 The Similarities Between These Events

What are the similarities between these three events, occurring in three different industries and separated by four decades? One event, what we might call an incident, resulted in mission delays and (ultimately) considerable embarrassment. Another, what we would term an accident, resulted in the loss of two lives, the injuries of many, and hundreds of millions of dollars in a lost aircraft and legal claims arising from the event. The Chernobyl event, widely described as a catastrophe, killed several people immediately, more over the decades that followed, led to the permanent abandonment of a small city, the construction of a concrete sarcophagus around the doomed property, and the pollution of huge swaths of previously productive farmland in several countries.

The underlying factor behind these three events is the failure by the operators to understand how the automated system worked, and their inability to take over operational control of the system when the automation needed a hand-off or showed signs of failing.

2.5 The Operational Environment in These Three Industries Compared to Automated Vehicles

In our three selected industries:

- The equipment being operated is all of a specific type, (e.g. Airbus 330 or Boeing 777). The operator is "type-rated" and operates only the specific system for which he or she has been trained…
  - But the automobile may be any of dozens of brands and hundreds of models, and other vehicles on the road may be 20 or more years old and may well be poorly maintained.

- The equipment being operated is maintained rigorously…
  - But, although some U.S. states have minimal vehicle maintenance requirements and periodic vehicle inspections, many, including the largest, have none.

- The time scale of unfolding events demanding attention may be minutes or hours…
  - But drivers have at most a few seconds to address an impending crash.

- There are comprehensive operating manuals that cover both normal and abnormal operations – manuals that must be read and understood in order to perform the required operations…
  - But even the once ubiquitous owner's manual is no longer made available to drivers; it has been replaced by online documentation that may or may not be reviewed. And there is no requirement that the operator possess any familiarity with vehicle operating procedures before taking the wheel.

- The software in aviation, aerospace, and nuclear power is typically quite stable over time, and when changes are made, operator retraining is performed prior to the update being placed into service…
  - In automobiles, software updates may occur whenever the manufacturer deems it appropriate (an approach followed, for example, by Tesla), and there is little if any concomitant operator training, thus adding to the likelihood of some unexpected outcome or loss of system reliability.

- Operators are trained to avoid inattention to their tasks and distractions are typically prohibited. Crews of two or more personnel operate at all times, such that one member can compensate for another who may be distracted or inattentive.
  - Automobiles are typically driven by a solo driver, who may be distracted by in-vehicle infotainment or devices (such as mobile phones) brought into the vehicle. Manufacturers paint a picture of the future driver relaxing with a magazine or television while the autonomous vehicle is in complete control.

2.6 The Capability and Preparation of the Operators in These Industries Compared to Those of Vehicle Operators

In our three chosen industries, operators:

- Are highly trained, for both normal and abnormal operating conditions...
  - o But automobile drivers, at least in the U.S., receive perfunctory training at best, and none for emergencies

- Are rigorously tested and licensed…
  - o But the driver's licensing process in the U.S. does not measure critical driving skills; and the license may be valid for five years or longer without any ongoing testing
.
- Follow specific procedures that cover both normal and off-normal operations…
  - o But automobile drivers follow no procedures while driving, save for the "rules of the road."

- Are medically examined regularly, and must be medically fit to maintain licensure…
  - o But most drivers in the U.S. are given a standard eye test that measures only static visual acuity and must meet little or no continuing medical standards.

- Must demonstrate proficiency in a provisional capacity at the hands of a senior instructor before being permitted to operate…
  - o But the provisional ("Graduated") license is generally overseen by parents, not experts, and it relates more to time behind the wheel than it does proficiency. There is typically no required proficiency demonstration for unusual or emergency events.

- Undergo periodic retraining and retesting…
  - o But for drivers in most of the U.S., no retraining or retesting is required, except (in some States) for drivers over a certain minimum age.

- May not work if they are under the influence of drugs or alcohol...
  - o But in the U.S., the BAC limit for driving is 0.08 percent; and there is no specified limit (or test) for drugs. Little random testing is done, and no regular testing.

(Note that, in the U.S., the operating environment and operator readiness are considerably more rigorous for interstate truck and bus drivers than they are for automobile drivers).

3. The Capability and Readiness of Drivers to Assume Control

Several authors have addressed some of the anticipated difficulties with human takeover of failed or compromised vehicle automation but have generally done so in the abstract. The present paper asks the question: by what hubris do we continue to design vehicles with advanced automation without accounting for the manner in which the human will interact with such automation when it fails or hands-over control, when extensive data from other industries (particularly aviation) highlights the often-flawed manner in which humans interact with technology in those industries, and therefore calls into question our assumptions for safe operations in the highway environment?

While it has been argued [9, 10] that drivers should have a deep understanding of how automated systems work in order to successfully respond when they fail, this goal seems all but unattainable in the automotive world when it has been shown to fail in other industries where training is rigorous, in depth, and continuous. While it is true that nuclear power plant operators as well as pilots and astronauts are thoroughly and repeatedly trained to have such underlying knowledge of the systems they operate, we do not see how such deep learning can be imparted to automobile drivers – given the time and resources required, the lack of a legal framework to require such training, and the competitive nature of the automobile industry in which manufacturers are loath to share information about their technical systems.

Stanton, writing in [9] describes the "utopian vision of the motor vehicle" that has an "onboard auto-driver, similar to the autopilot in aircraft (to) take over the driving tasks, allowing the human driver to work, rest or play." He opines that "the Catch-22 of vehicle automation is that, while car owners are stripped of the need to perform driving tasks, they are still required to monitor their auto-driver and take manual control if the situation demands. However, when vehicles become fully autonomous, even the most observant human driver's attention will begin to wane. Their mind will begin to wander, and they may start to mentally switch off from the job of driving.". As Stanton and others paint this "utopian vision," they typically include the image of the driver being able to engage in other activities or, simply, rest. These "utopian" ideals, which always include such distractions, exacerbate the conflict between a proposed need for deeper understanding of system operation and loss of focus on the driving task, should takeover from automation become necessary. While this is most commonly addressed in discussions about fully autonomous vehicles, it is of particular concern with Level 3 and Level 4 systems.

Stanton's simulator and test-track research has shown that drivers of automated vehicles are generally less effective in emergencies than drivers of manual vehicles, and he has "repeatedly witnessed the failure of drivers to intervene when systems fail whereas almost all drivers of manual vehicles recover in the same situation."

As a result of his research, Stanton has suggested that automation must have graduated, gradual hand-over if it is to successfully support human drivers. And he proposes that the interface between the driver and the vehicle automation be in the form of a "chatty co-pilot, not a silent auto-pilot."

Nunes, Reimer and Coughlin [10] strike a similar tone. They believe that one approach to this problem is to educate consumers about how the automated system works, and to alert them to safety concerns that may arise. Yet, they point out, "self-driving cars are underpinned by sophisticated technologies that are hard to explain or understand." (p. 170). They believe that "developers are designing such products to be easy to use. … However, users are then less able to anticipate how the underlying systems work, or to recognize problems and fix them." (p. 170)

Setting a rather different tone than many other writers, these authors believe that some form of human intervention will always be required, regardless of the degree of automation. The irony of this statement comes about from the same authors' admission that governments worldwide are freeing developers of automated vehicles from having to meet current safety requirements such as providing a steering wheel, rear view mirror, and manual braking control.

Other ironies exist. If we accept the premise that autonomous vehicles will always require some degree of user intervention, then individuals with cognitive impairments or age related cognitive decline may find the operation of such vehicles challenging. Yet these are cohorts that are expected to be among the greatest beneficiaries of automated vehicles.

Further, existing legislation in the U.S. makes no mention of either competency requirements or proficiency testing for users, and, without such standards, these authors worry, the risk of incidents might increase.

The report ends with a call to policymakers to recognize that "driverless does not, and should not, mean without a human operator;" and that automation (essentially) changes the work that people must perform – it does not eliminate it. They further posit that vehicle operators should be required to demonstrate competence – "that proficiency standards are necessary for users of autonomous vehicles and that competency should be tested by licensing authorities and should supplement existing driving permits." (p. 171). They further advocate mandatory regular checks on user competency "so that proficiency is kept up as cognitive abilities change, and technology evolves." (p. 171) This is a laudable and appropriate position, but, as discussed herein, likely impossible to achieve.

In her seminal chapter, “Ironies of Automation,” Bainbridge [11] could be writing for those responsible for autonomous vehicles. She describes, for example, two ironies stemming from “the designer’s view … that the operator is unreliable and inefficient, so should be eliminated from the system.” (p. 272). The first irony is that “designer errors can be a major source of operating problems,” just as we have seen with, for example, the problematic algorithm that led to the false positive situational interpretation that resulted in a pedestrian death in a crash with an Uber vehicle in Tempe, Arizona [12]. The second irony is that “the designer who tries to eliminate the operator still leaves the operator to do the tasks which the designer cannot think how to automate.” Compare this expressed irony to the Cunningham and Regan [13] and Wolmar [4] examples of autonomous vehicle failures under conditions of snow, dust, or even rain covered roads, hand-signalling by police officers, or roadside construction zone detours and sudden lane changes and drops. Bainbridge’s prescient writing reminds us that skills deteriorate when they are not used, and so an erstwhile experienced operator may become an inexperienced one when suddenly having to take over for a failed automated process that has functioned properly for an extended period. She argues that, “when manual takeover is needed there is likely to be something wrong with the process, and the operator needs to be more rather than less skilled to handle it.” (p. 272). Both Cunningham and Regan [13] and Nunes, Reimer and Coughlin [10] suggest that, in order to properly be prepared to take over in the event of automation hand-off or failure, the operator of an autonomous vehicle needs to have a deep understanding of system operation. Perhaps the “safety driver” who was “unable to prevent” the pedestrian fatality in Temple, Arizona would have been more successful had he or she possessed such deep knowledge, sufficient to timely override the faulty decision-making algorithm within the Uber vehicle’s software. Here, too, Bainbridge has offered cogent arguments some 30 years before the fact, and summarizes with the rather pessimistic view that the “current generation of automated systems” which are monitored by “former manual operators” are riding on the learned skill sets of these operators, and that future generations may not possess such skills, a view that could well apply to tomorrow’s safety drivers. Promised distractions from the driving task will further exacerbate this issue.

Eriksson and Stanton [14] state: “When the driver is assumed to resume control of a vehicle when its operational limits are reached, a critical weakness in the system is exposed. As the driver (has) been out of the control loop for an extended period of time, they may be a victim of some of the ironies of automation, where situation awareness is reduced.” Under such circumstances, they posit, the driver must receive support and guidance necessary to re-enter the control loop – and they propose the paradigm of the “chatty co-driver.” In their view, this facet of automation would provide continuous feedback via specialized user interfaces following the convention of the Gricean Maxims of successful conversation [15].

4. Two Approaches to Driver Preparation

It does not seem likely that, in the future, prior to the introduction of Level 5 vehicles into the traffic stream, either the time scale of motor vehicle operations or the physical roadway spacing in which such vehicles operate will change, except for an increase in the density of both, nor that the competition between vehicle manufacturers will permit designs or implementation of automated systems in vehicles to be harmonized. Therefore, the best hope for reducing the potential for errors when automation fails or requires a handoff lies with the human operator. And since we are not likely to see, at least in the U.S., greater rigor in the medical fitness arena or in the testing phase of the driver licensure process, it seems undeniable that improvements will have to come in the realm of driver training and preparation for dealing with automation, or in the constant feedback provided by an interface to equip the operator with current knowledge of system status and function. Either of these two approaches would mark a major step forward, although neither is likely to receive Government support or enforcement.

Although it has been shown that, “even brief training in how to respond to AV failure seems promising [13], In the U.S., at least, it can be argued that driver training has not advanced in recent years – if anything, such preparation to drive has been declining over time, with the exception of certain States’ Graduated Driver Licensing (GDL) programs. Some authors have suggested that a new era of driver training is necessary, with potential vehicle purchasers required to receive training in vehicle showrooms as part of the new car purchasing experience, and we agree that specialized training and rehearsal of a driver’s interaction with vehicle automation would be useful if we are to close the gap, to even a small degree, between vehicle drivers and those who operate nuclear power plants, aircraft or space vehicles. Others, including Eriksson and Stanton [14], recommend the “chatty co-driver” approach, and this novel intervention also seems to have potential. There is, however, no existing model of such a system in commercial use, and the closest approximation would appear to be the currently available on-line owner’s manual. Such manuals are not, of course, real time information systems, and they require the operator to seek them out and investigate them thoroughly for them to be at all effective.

In short, two theories have emerged that purport to address a means to fill the gap when a distracted or inattentive driver is confronted with a (potentially) sudden need to re-enter the control loop and take manual control of the vehicle in the event of automation failure or hand-off. These two approaches involve specialized training in the workings and failure modes of the automation; and a continuously informative user interface to keep the driver abreast of the status and functioning of the automation at all times. Each is intended to fulfil the goal of preparing the driver to take over control at a moment’s notice when it becomes necessary if the automation can no longer manage the vehicle’s movements.

There are, of course, potentially serious disadvantages to each approach. In the first case, training to a level presumably necessary to handle such automation failures or hand-offs is nearly impossible given the size of the driving population, the uniqueness of each manufacturer’s automation implementation, and the logistics of requiring the purchasers of automated vehicles to participate in such training. Further, training to a level necessary to respond to any and all failures or hand-offs (especially when many may not be known) as is the case with pilots, astronauts, and nuclear power plant operators (who *still* exhibit occasionally fatal misunderstandings of the automation), is an unreasonable and unreachable expectation given the nature of the driving environment and driver availability to participate in such training. In addition, recurrent training, to refresh skills or keep pace with changes in automation, routine in these other industries, is less feasible still. In the case of the “chatty co-driver,” such interfaces would have to be designed for every implementation of vehicle automation, and system designers would be tasked with designing such a supportive interface for hand-off and failure modes that might not be fully understood. On the implementation side, a near-constant source of voice communication might provide exactly the type of in-vehicle information that drivers of automated vehicles are hoping to escape – seeking rest and relaxation (read inattention and distraction) while the vehicle drives itself. Thus, there is the risk that drivers will turn off (if possible) the interface or learn to ignore it, thus defeating its very purpose.

Nonetheless, given the constraints of the operating environment and the overall lack of preparation of vehicle operators to take over from an automated system, these two approaches seem to offer promise to improve the likelihood of success in such takeovers.

A failure to begin to evaluate interventions such as these would be abrogating our responsibility to maximize road safety in Level

3 and, especially Level 4 vehicles. It is with hubris that we continue to move forward with automation technologies while failing to prepare present and future vehicle operators to interact with those technologies, particularly when they require human intervention; and such automated systems are likely to require such intervention for many years to come.

Accordingly, we have proposed a thought experiment to examine the feasibility of the two interventions discussed above.

5. A Thought Experiment

Other than those authors who seem to think that the movement toward fully autonomous vehicles will provide a utopian, highly functional and risk-free driving environment, others have pointed out that driver complacency, distraction and inattention, coupled with a lack of understanding of the inner-workings of the automated system, will result in a dangerous driving environment for years to come. Although the data set is small and the results, therefore, not significant, crashes per million vehicle miles are far higher in autonomous vehicles on the streets than they are for the overall vehicle population [16], and the number of handoffs of the automation to a "safety driver" (called "disengagements") are, not surprisingly, quite high. *Recode* has reported on 2017 data showing that Uber disengagements occurred nearly once per driven mile, and that "critical" disengagements (to avoid hitting a person or causing more than $5,000 in property damage) occurred, on average, once per 125 miles driven [17]. It has been suggested that a form of specialized operator training, or an in-vehicle interactive assistant could enable a reduction in otherwise foreseen driver failures to timely respond to automation failures or hand-offs.

We, therefore, propose a thought experiment to examine the viability of these two possible interventions.

5.1. A Training Protocol

A training protocol, likely an interactive, computerized series of lessons based on existing online operators' manuals, would be developed. For testing purposes, this protocol would be limited to a specific, challenging subset of possible automation failures or handoffs. In order to be acceptable to the automotive industry and the public alike, a pilot test of the effectiveness of the protocol would likely have to be conducted in automotive dealerships with volunteer participants and/or as part of the vehicle purchasing process. A reward would be provided for participation, perhaps in the form of dealership merchandise. Participants would be encouraged to bring to the session their own choice of entertainment or relaxation (e.g. music, reading materials, computer games, etc.). A 20-30-minute session conducted in a part-task interactive driving simulator in the showroom would first familiarize the participants with the selected subset of automated features and the failure and recovery modes for these features, provide an opportunity for any questions that the participant may have about the training, conduct the actual training, and then test its effectiveness on simulator driving scenarios. Scenarios that are functionally equivalent to those used in the training session would be used to test the appropriateness and timeliness of the participant's responses. Ideally, a follow-up session would test retention of the information after several weeks. Participants would be queried regarding their opinion of, and satisfaction with, the training model. A careful review of failures would need to be kept to advise on possible revisions to the protocol.

5.2. A Smart Assistant

The intervention of a "smart assistant" or "chatty co-driver" would also be introduced through part-task simulation, and would follow the same protocol discussed above but, since this system is meant to operate on the road in real time, upfront training would be limited to an introductory familiarization session on how the system functions, and how it should be used. After this introduction, any participants' questions would be addressed.

The same scenarios and automation failures/hand-offs as in the training model would be presented, with the (previously developed) smart assistant providing continuous information and feedback to the participant/driver. The same equivalent form simulator scenarios as used in the training protocol would be used here, again to test the appropriateness and timeliness of the participant's response to failures and hand-offs. Again, a follow-up session would test retention of the information after several weeks. And again, participants would be queried regarding their opinion of, and satisfaction with, the smart assistant system. Finally, as in the training protocol, a review of failures would be critical for designing any necessary revisions to the smart assistant.

It is suggested that this series of trials would shed light on the functionality, viability, and consumer (and manufacturer and regulator) acceptance of the training approach vs. the smart assistant.

6. Conclusions

We have described three different incidents that have occurred in three different industries, in each case where the operators were highly trained, rigorously tested, and medically fit. We pointed out the vast differences between the operating environment and operator readiness in these industries compared to that in the highway setting. We then explored different theories of the "ironies of automation," and looked at different approaches to addressing the safety implications of these ironies, particularly for Level 3 and 4 automation. We proposed a thought experiment to evaluate two such approaches – in-depth training into system operation and failure, and an on-board "chatty assistant" to keep the driver continuously informed about automation state. We express our concern that, given the experiences with automation in highly regulated and controlled industries and the current lack of such regulations and control in the automotive field, it is with considerable hubris that we continue to advance vehicle automation with full knowledge that driver takeover will be required for many years to come but without any real commitment to driver preparation for such takeover.

7. References

[1] 'Hyatt, K., Paukert, C.: 'Self-driving cars: A level-by-level explainer of autonomous vehicles', https://www.cnet.com/roadshow/news/self-driving-car-guide-autonomous-explanation, accessed 15 May 2018

[2] 'Automated vehicles for safety', https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety, accessed 15 May 2018

[3] Godsmark, P. 'The definitive guide to the levels of automation for driverless cars', https://driverless.wonderhowto.com/news/definitive-guide-levels-automation-for-driverless-cars-0176009/, accessed 15 May 2018

[4] Wolmar, C.: Driverless cars: 'On a road to nowhere' (London Publishing Partnership, 2018)

[5] National Aeronautics and Space Administration, 'Apollo 10 Mission Report' (Houston: Manned Spacecraft Center, 1969), pp. 1-323

[6] 'The Chernobyl Gallery', http://chernobylgallery.com/chernobyl-disaster/cause/, accessed 15 May 2018

[7] 'Board meeting: Crash of Asiana flight 214 accident report summary' (National Transportation Safety Board, June 24, 2014, https://www.ntsb.gov/news/events/Pages/2014_Asiana_BMG-Abstract.aspx, accessed 15 May 2018

[8] Mumaw, R.J. **'**Addressing mode confusion using an interpreter display' Contractor Technical Report, NASA Contract # NNL16AA15C. San Jose State University Research Foundation: Moffett Field, CA. DOI 10.13140/RG.2.2.27980.92801, 2018

[9] Poulin, C., Stanton, N.A., Cebon, D.: 'Response to: Autonomous vehicles', Ingenia online, 2015, Issue 62, unpaginated,

http://www.ingenia.org.uk/Ingenia/Articles/943, accessed 17 May 2018

[10] Nunes, A., Reimer, B., Coughlin, J.F.: 'People must retain control of autonomous vehicles', Nature, 2018, 556, pp. 169-171

[11] Bainbridge, L. 'Ironies of automation', in Rasmussen, J., Duncan, K., Leplat, J. (Eds.): 'New Technology and Human Error' (Wiley, 1987) pp. 271-283

[12] Efrati, A. 'Uber finds deadly accident likely caused by software set to ignore objects on road', 7 May, 2018, https://www.theinformation.com/articles/uber-finds-deadly-accident-likely-caused-by-software-set-to-ignore-objects-on-road, accessed 14 May 2018

[13] Cunningham, M.L., Regan, M.A. 'Driver distraction and inattention in the realm of automated driving', IET Intell. Transp. Syst., 2017, pp. 1-7

[14] Eriksson, A., Stanton, N.A. 'The chatty co-driver: A linguistics approach to human-automation-interaction', Human Factors in Organizational Design and Management – XI, Nordic Ergonomics Society Annual Conference – 46, Undated, unpaginated

[15] Grice, H.: 'Logic and conversation', in Cole, P. and Morgan, J. L. (Eds.): 'Speech Acts' (New York, Academic Press), 1975), pp. 41-58

[16] Els, P. 'How much testing will prove automated cars are safe?', Automotive IQ, 2018, Unpaginated, https://www.automotive-iq.com/autonomous-drive/articles/how-much-testing-will-prove-automated-cars-are-safe, accessed 15 May 2018

[17] Bhuiyan, J. 'Uber's autonomous cars drove 20,354 miles and had to be taken over at every mile, according to documents', Recode, 2017, https://www.recode.net/2017/3/16/14938116/uber-travis-kalanick-self-driving-internal-metrics-slow-progress, accessed 16 May 2018

Under which driving contexts do drivers decide to engage in mobile phone related tasks? An analysis of European naturalistic driving data

Tina Morgenstern [1*], Frederik Naujoks[2], Josef F. Krems [1], Andreas Keinath [2]

[1] Cognitive & Engineering Psychology, Technische Universität Chemnitz, Chemnitz, Germany
[2] BMW Group, München, Germany
[*]tina.morgenstern@psychologie.tu-chemnitz.de

Abstract: Mobile phone related task engagement while driving has increased dramatically over the past years. However, research has shown that drivers attempt to compensate for the associated performance degradation in the primary driving task by using various self-regulatory strategies, such as deciding *when* to engage in a secondary task. Unfortunately, there are only a few existing studies that focus on contextual factors associated with secondary task initiation. Goal of the present study was to investigate which driving contexts encourage drivers to initiate a mobile phone related task using European naturalistic driving data. In total, 165 trip segments involving mobile phone engagement were analysed. The driving context at the moment of task initiation was compared to the context 30 seconds prior to task initiation. With the exception of conversation, the results show that drivers were much more likely to be stopped at task initiation than 30 seconds prior, indicating that most drivers stopped their vehicle before initiating the secondary task. Further, for texting or browsing tasks, making turns or driving in a stable traffic flow was significantly less likely at task initiation. The results suggest that drivers choose to engage in mobile phone tasks when the driving task demand is low.

1. Introduction

In recent years, the use of mobile phones while driving has increased tremendously [1, 2], particularly among younger drivers [3, 4]. However, mobile phone interaction while driving, especially texting, can adversely affect driving performance. Texting can cause slower reaction times [5, 6] and more lane deviations [7, 8]. Previous studies also show an alarmingly high crash risk of texting compared to other common secondary tasks (e.g., eating and drinking, talking with passengers) while driving [9, 10].

At the same time, there is evidence from simulator studies that drivers use self-regulatory strategies on an *operational* level to decrease the driving demand during secondary task engagement, such as limiting the number of lane changes [11], increasing the following distance to a lead vehicle [5, 12] or reducing speed [6, 8, 13]. The effect of speed reduction during secondary task engagement is a particularly common finding reported across different driving simulator studies. However, analyses of naturalistic driving data showed that these effects are rather small, if they are found at all. For instance, Schneidereit, Petzoldt, Keinath et al. [14] examined data from the SHRP 2 naturalistic driving study and found only a small indication regarding a speed adjustment for texting while driving. The engagement in other secondary tasks, such as smoking or eating, did not significantly alter speed. Tivesten and Dozza [15] found comparable results when analysing visual-manual phone task engagement in their Swedish naturalistic driving study, revealing little to no changes in speed prior to or after mobile phone task initiation.

Based on these results, it seems more likely that drivers self-regulate on a *strategic* level, such as deciding *when* to engage in a secondary task while driving. Some evidence exists that drivers engage in secondary tasks more frequently when the driving task demand is low, for example during slow speeds [16] or when stopped [4, 15, 17]. Huisingh, Griffin and McGwin Jr. [18] found in their roadside observational study that drivers were engaged in secondary tasks much more often when the car was stopped. However, when focusing on mobile phone calls, the prevalence of this secondary task did not differ significantly depending on the vehicle speed. For texting or dialing tasks it even turned out that more drivers actually dialed or texted at speeds greater than 50 mph than at lower speeds or while stopped [18]. These results are somewhat surprising as visual-manual mobile phone tasks, such as texting, are considered as one of the most distractive and dangerous secondary tasks to engage in while driving that force drivers to take their eyes off the road, which, in turn, lead to a high safety-critical risk [9].

Aside from some of the findings regarding speed, evidence also exists that drivers' secondary task engagement is associated with specific road types. Huisingh, Griffin and McGwin Jr. [18] noted that the overall secondary task engagement was more common on local than on arterial roads; however, texting and dialing tasks actually occurred more frequently on arterial roads (i.e., in urban centres). Further, there are indications that drivers tended to avoid secondary task engagement in dense traffic environments, while turning, or under adverse lighting or weather conditions [3, 15].

In fact, it appears that there are certain contextual factors where drivers are either more or less likely to engage in secondary tasks. However, findings on mobile phone related tasks are rather inconsistent regarding the actual extent of drivers' behaviour adaptation to different driving contexts. For instance, some studies have shown that drivers' mobile phone engagement was much higher when the vehicle was stopped (e.g., at a red light) [4, 19], whereas in other studies the exact opposite was found [18]. Furthermore, most of the previous findings are based on roadway observations or survey studies. Only a few studies on this topic currently exist that use naturalistic driving data [15, 19]. Naturalistic driving data create a clear image of drivers' mobile phone behaviour across different driving contexts. Moreover, it allows for the comparison of the driving context at the precise moment of task initiation with the driving context before the mobile phone task was initiated. Thus, contextual factors increasing the prevalence of mobile phone task initiation can be assessed. The aim of the present study was therefore to identify the contexts under which drivers decide to engage in mobile phone related tasks using European naturalistic driving data.

2. Method

The current study is based on European naturalistic driving data collected in the UDRIVE project [20]. Within UDRIVE 120 cars in five countries (France, Germany, Poland, United Kingdom, and Netherlands) were equipped with seven video cameras (three forward, one cabin, one cockpit, one face and one footage camera) and a data acquisition system that was developed for the project (e.g., to record GPS, speed behaviour, brake pressure or steering wheel angle). Drivers' natural behaviour was observed for up to two years. Overall, 192 drivers participated in the study [21].

2.1. Sampling and Annotation

The analyses presented in this paper rely on a dataset containing four randomly selected trips per driver. For our analyses we used all trip segments in which a mobile phone interaction took place. The trip segments were annotated using video data regarding the *main mobile phone related task* (i.e., conversation hand-held, conversation hands-free, texting/ browsing, reading hand-held, reading hands-free, holding, other; for a detailed description of the tasks see Table 7, Appendix A), the precise moment of *task initiation* and the precise moment of *task conclusion*. Task initiation and conclusion were defined as the first/ last glance or hand movement (whatever occurred first/ last) towards the mobile phone. At task initiation (further referred to as "I-0") we also annotated if *other passengers* were present (i.e., yes, no) as well as *weather* (i.e., clear, rain, snow, fog, other) and *lighting conditions* (i.e., daylight, dawn/ dusk, darkness). *Locality* (i.e., urban-residential, urban-motorway, rural, motorway/ highway, other), *traffic density* (i.e., free flow, free flow with restriction, stable flow, unstable flow, traffic jam/ stop-and-go, other), *stopping* (i.e., yes, no), *location when stopped* (i.e., traffic light, traffic sign, parking lot, traffic jam, other) and *turning* (i.e., yes, no) were annotated at I-0 and also 30 seconds prior to task initiation (further on referred to as "I-30").

Overall, 305 trip segments were annotated. 269 of these trip segments were relevant, i.e. contained a clear mobile phone related task (in some cases, for example, it was not obvious whether the driver engaged in a hands-free mobile phone conversation or talked with a passenger). The 269 trip segments stemmed from 129 different trips. For further analyses, we randomly selected one segment per trip in case multiple trip segments per task category stemmed from one trip. This was done to avoid an overrepresentation of single trips. Thus, 104 trip segments were excluded from the analyses (see Table 1).

Table 1 Frequencies across mobile phone tasks for the dataset including all trip segments and the dataset with one segment per trip

Task category	Dataset with all trip segments	Dataset with one segment per trip
Conversation hand-held	19	18
Conversation hands-free	7	6
Texting/ browsing	143	64
Reading hand-held	37	30
Reading hands-free	8	8
Holding	21	16
Other	34	23
All	**269**	**165**

2.2. Sample Description

The 165 trip segments analysed consisted of 57 different drivers (30 female, 27 male) with a mean age of 40 years (*SD* = 11.25). Most of the drivers in our sample came from Poland, whereas the fewest originated from Germany. Table 2 gives an overview of the sample characteristics.

Table 2 Sample description per secondary task category

Mobile phone task category	Number of trip segments	Number of drivers	Gender		Mean Age (*SD*)	Operational Site				
			Female	Male		DE	FR	NL	PL	UK
Conversation										
hand-held	18	15	7	8	40.75 (11.38)	2	2	0	9	2
hands-free	6	6	3	3	37.00 (9.03)	0	2	1	2	1
Texting/ browsing	64	34	17	17	37.59 (11.26)	1	10	3	12	8
Reading										
hand-held	30	22	14	8	36.35 (10.04)	2	5	1	6	8
hands-free	8	6	4	2	38.4 (7.02)	0	1	1	3	1
Holding	16	14	8	6	36.08 (8.65)	1	3	1	4	5
Other	23	19	10	9	38.07 (11.97)	1	6	0	7	5
All	**165**	**57**	**30**	**27**	**40.08 (11.40)**	**4**	**14**	**5**	**18**	**16**

Especially for texting or browsing, it must be noted that multiple trip segments stemmed from one driver. However, this was rarely the case for the other mobile phone related tasks (see Table 2).

2.3. Analyses

Prevalence ratios were used to assess the association between the frequency of different contextual factors and the initiation of mobile phone related tasks. Prevalence ratios are calculated exactly like risk ratios and indicate how common an event is in one group or data collection point relative to another group or data collection point [22]. More precisely, the proportion of specific contextual factors (e.g., a free flow traffic condition) at I-0 (i.e., at task initiation) was divided by the proportion of the same contextual factors at I-30 (i.e., 30 seconds prior to task initiation). A prevalence ratio less (greater) than 1 means that the prevalence for the respective contextual factor at I-0 is lower (higher) than at I-30. Prevalence ratios were calculated for locality, traffic density, stopping and turning. Frequencies are reported for the other categories.

3. Results

During most annotated trips, the drivers engaged in texting or browsing, followed by hand-held reading and other mobile phone related tasks (see Table 1). Only 18 observed trips included hand-held mobile phone conversations. In further analyses, “conversation hand-held” and “conversation hands-free”, “reading hand-held” and “reading hands-free”, as well as “holding” and “other” were combined due to their low number of observed events.

3.1. Texting/ Browsing

Texting or browsing tasks were the most observed mobile phone related tasks in our sample. The mean duration of texting or browsing was 46 seconds (*SD* = 50.78), ranging from 3 to 271 seconds. Other passengers were present in 16% of the trip segments. Most of the trip segments in which drivers engaged in texting or browsing took place in daylight (78%), under clear weather conditions (93%) and in an urban area (68%).

The prevalence ratios indicate an association between the initiation of texting or browsing and traffic density, stopping and turning (see Table 3). Specifically, the data show that a stable traffic flow was observed significantly less often at I-0 than at I-30. In contrast, the prevalence of the “other traffic density” category was two times higher at I-0 than at I-30. This category contains all events in which the vehicle was stopped (e.g., at a red light) and therefore traffic density could not be assessed. This is also reflected in the high prevalence ratio of stopping, indicating that the prevalence of a stopped vehicle at I-0 was 3.5 times higher than at I-30. Furthermore, we found a significant prevalence ratio regarding turning, such that turning occurred less often at I-0 in comparison to I-30.

Table 3 Prevalence ratios and 95th confidence intervals regarding locality, traffic density, stopping and turning for texting or browsing tasks

Contextual factor	Prevalence ratio	95th CI
Locality		
Urban residential	0.92	0.67-1.25
Urban motorway	0.92	0.34-2.45
Rural	1.60	0.50-5.21
Motorway	0.57	0.20-1.66
Other	2.45	0.68-8.79
Traffic density		
Free flow	0.55	0.26-1.17
Free flow with restriction	0.55	0.14-2.21
Stable flow	0.37*	0.17-0.78
Unstable flow	0.92	0.13-6.32
Traffic jam	1.38	0.52-3.64
Other	3.17*	1.69-12.71
Stopping	3.58*	1.95-5.93
Turning	0.20*	0.05-0.91

Note. *Significant prevalence ratios (i.e., 95th CI does not cross 1).

Regarding vehicles' stopping location at time I-0, it could be shown that in more than half of the events the vehicle stopped at a (red) traffic light, whereas waiting at a traffic sign (e.g., a stop sign) was not observed in our sample (see Fig.1).

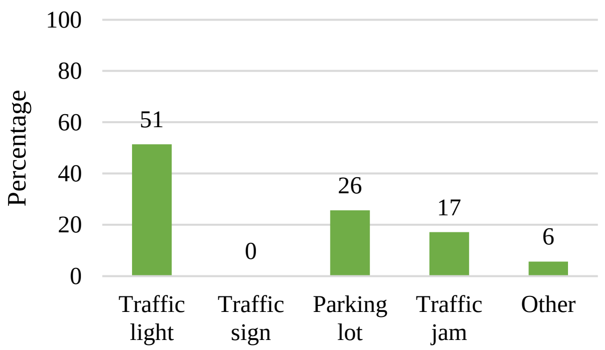

Fig. 1*. Percentage of annotated stopping locations when initiating texting or browsing tasks*

3.2. Conversation

Hand-held or hands-free mobile phone conversations lasted on average 250 seconds (*SD* = 407.34), ranging from 35 to 1464 seconds. Other passengers were only present in 8% of all events. Here again, most trips including a mobile phone conversation occurred in daylight (75%), under clear weather conditions (86%) and in an urban area (64%).

For mobile phone conversation, no significant associations existed between the initiation of a conversation and specific contextual factors (see Table 4). A stopped vehicle was more frequently observed at I-0 than at I-30; however, this effect was not statistically significant.

Table 4 Prevalence ratios and 95th confidence intervals regarding locality, traffic density, stopping and turning for conversation tasks

Contextual factor	Prevalence ratio	95th CI
Locality		
Urban residential	1.24	0.71-2.15
Urban motorway	0.48	0.05-4.90
Rural	0.64	0.12-3.45
Motorway	0.95	0.27-3.34
Other	0.95	0.15-6.19
Traffic density		
Free flow	0.64	0.27-1.50
Free flow with restriction	1.91	0.39-9.32
Stable flow	0.41	0.12-1.40
Unstable flow	–	–
Traffic jam	–	–
Other	2.55	0.79-8.17
Stopping	3.82	0.93-15.63
Turning	0.76	0.24-2.48

Note. "–"Prevalence ratios could not be calculated due to missing values in this category.

Due to the small sample size of mobile phone conversation events, vehicles' location when stopping at I-0 will not be reported.

3.3. Reading

Mobile phone tasks involving reading a message/ post (hand-held or hands-free) lasted on average 18 seconds (*SD* = 12.54), ranging from 1 to 61 seconds. Another passenger was present in 24% of all events. Further, reading was mostly observed in daylight (74%), under clear weather conditions (94%) and in an urban area (75%).

The prevalence ratios indicate a link between the initiation of reading a message/ post on the mobile phone and the category "other traffic density" as well as the category "stopping" (see Table 5). The prevalence of "other traffic density" was five and the prevalence of a stopped vehicle was two times higher at I-0 than at I-30.

Regarding vehicles' stopping location at time I-0, most events occurred when the vehicle was stopped at a (red) traffic light, followed by stopping in a traffic jam (see Figure 2).

Table 5 Prevalence ratios and 95th confidence intervals regarding locality, traffic density, stopping and turning for reading tasks

Contextual factor	Prevalence ratio	95th CI
Locality		
Urban residential	1	0.71-1.40
Urban motorway	0.92	0.20-4.24
Rural	0.92	0.20-4.24
Motorway	1.15	0.34-3.92
Other	0.92	0.06-14.10
Traffic density		
Free flow	0.58	0.25-1.34
Free flow with restriction	0.69	0.17-2.86
Stable flow	0.55	0.14-2.14
Unstable flow	0.37	0.08-1.77
Traffic jam	1.53	0.62-3.73
Other	5.04*	1.21-20.89
Stopping	2.29*	1.21-4.33
Turning	0.46	0.12-1.70

Note. *Significant prevalence ratios (i.e., 95th CI does not cross 1).

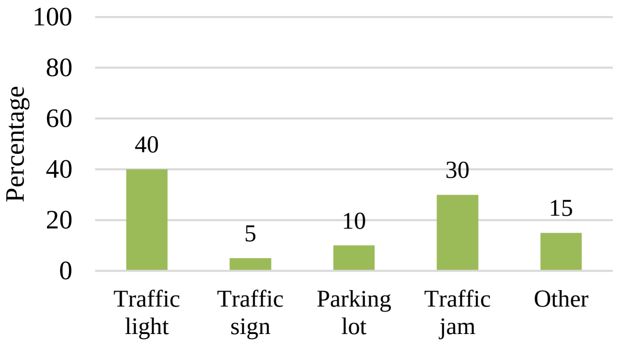

Fig. 2. *Percentage of annotated stopping locations when initiating reading tasks*

3.4. Other Mobile Phone Related Tasks

Other mobile phone related tasks include, for example, holding the phone or taking a picture. On average, these tasks lasted around 26 seconds (*SD* = 51.13), ranging from 1 to 292 seconds. Another passenger was present in 21% of these events. Other mobile phone related tasks mainly occurred in daylight (82%), under clear weather conditions (92%) and in an urban area (65%).

The prevalence ratios were statistically significant for the contextual factors “other traffic density” and “stopping” (see Table 6). The prevalence of “other traffic density” was nearly ten times higher at I-0 than at I-30. This is also reflected in the significant prevalence ratio for “stopping”, indicating that a stationary vehicle was much more common at I-0 than at I-30. A statistically significant association was not found for the other contextual factors.

Table 6 Prevalence ratios and 95th confidence intervals regarding locality, traffic density, stopping and turning for other mobile phone related tasks

Contextual factor	Prevalence ratio	95th CI
Locality		
Urban residential	1.05	0.67-1.64
Urban motorway	0.89	0.24-3.30
Rural	1.02	0.41-2.52
Motorway	0.89	0.19-4.14
Other	1.22	0.19-8.06
Traffic density		
Free flow	0.77	0.43-1.40
Free flow with restriction	0.25	0.06-1.15
Stable flow	1.02	0.41-2.52
Unstable flow	–	–
Traffic jam	2.68	0.29-24.53
Other	9.81*	1.34-71.49
Stopping	4.46*	1.44-13.82
Turning	0.89	0.24-3.30

Note. *Significant prevalence ratios (i.e., 95th CI does not cross 1); “–“Prevalence ratios could not be calculated due to missing values in this category.

As shown before, most stops occurred at a (red) traffic light, followed by traffic jam (see Figure 3).

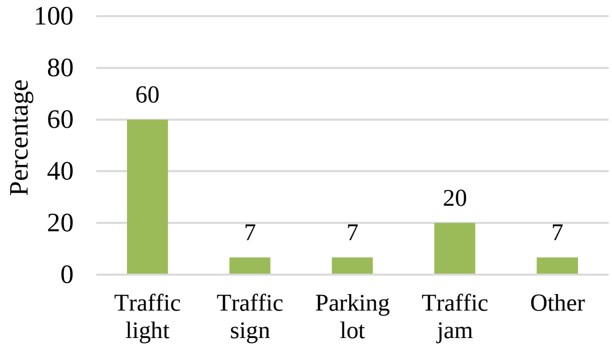

Fig. 3. *Percentage of annotated stopping locations when initiating other mobile phone related tasks*

4. Discussion and Conclusion

The aim of the present study was to investigate the contexts under which drivers engage in mobile phone related tasks using European naturalistic driving data. Prevalence ratios were calculated to assess the association between different contextual factors and the initiation of a specific mobile phone related task (i.e., texting or browsing, conversation, reading or another mobile phone related task). The results show a very clear pattern. The prevalence of a stopping vehicle was much higher at task initiation than 30 seconds prior to task initiation. This is in line with other study findings [4, 15]. Hence, drivers seem to selectively engage in mobile phone tasks when the driving task demand is low. Apart from mobile phone conversation, this effect existed for all other analysed mobile phone related tasks. One possible explanation why this effect was absent for mobile phone conversations is that these tasks include both incoming and outgoing calls. The drivers themselves initiate outgoing calls, whereas incoming calls are beyond the drivers' control. Although the driver has the choice to ignore the phone call, it can be suggested that in most cases drivers' curiosity (or need to know who calls) is too powerful.

Further, our analyses showed that drivers initiated texting or browsing tasks significantly less often when driving in a stable traffic flow or when turning. Such driving situations normally require much attention as traffic conditions can change rapidly, which might increase the driving task demand. Thus, drivers seem to avoid initiating texting or browsing in these complex situations. This corresponds to what was found by Tivesten and Dozza [15], showing that drivers initiate visual-manual mobile phone tasks more often *after* making sharp turns.

It must be pointed out that the results for stable traffic flow and turning were only significant for texting or browsing tasks. There was no association between task initiation and the presence (or absence) of these contextual factors for the other mobile phone related tasks. As texting requires visual, manual and cognitive resources, it is one of the most dangerous secondary tasks to conduct while driving. The meta-analysis by Caird, Johnston, Willness et al. [7] showed that texting while driving adversely impacts nearly all aspects of driving performance due to the repeated off-road glances necessary. Although reading text messages showed smaller effect sizes, driving performance was still negatively affected. Consequently, it can be assumed that with increasing secondary task difficulty, the more important contextual factors (e.g., traffic density, turning) become for secondary task initiation.

However, it is important to state that the sample sizes of the present study are rather small. Mobile phone conversations, for example, were only observed in 24 events, leading to a low level of statistical power. Analyses of larger sample sizes shall be performed to validate our findings. Moreover, in some cases multiple trip segments stemmed from one driver, which could lead to an overrepresentation of single drivers and thus might bias our results.

It has to be kept also in mind that in our analyses contextual factors *within* a single trip were compared. This was done to examine whether the traffic situation 30 seconds prior to drivers' engagement in a mobile phone related task differed from that at task initiation. This may have led the driver to consciously choose to (not) engage in the mobile phone related task at that precise moment. However, the influence of other contextual factors, such as passenger presence, cannot be investigated with this approach. For this, comparisons with baseline trips (i.e., trips without secondary task engagement) would be necessary. In general, although naturalistic driving data give insight into natural driving behaviour, it remains unclear why drivers act as they do. Personal motives and reasons are not directly apparent. Here, surveys and focus groups might provide additional information.

Nevertheless, our research gives an initial insight into drivers' self-regulatory behaviour adaptation on a strategic level when engaging in different mobile phone related tasks. The findings indicate that drivers strategically decide when to engage in a mobile phone related task by choosing low-demand driving situations. However, even though drivers across most of our analysed trip segments initiated the mobile phone task when stopped at a red light, there cannot be a presumption that this behaviour is "safe". For this, the moment of task conclusion must be further examined. If drivers, for example, continued with secondary task engagement after the light turns green, this poses a real traffic safety danger. Therefore, future studies focusing on specific contextual factors, such as red light situations, are necessary to better understand how often, why and how drivers use such situations for secondary task engagement.

5. Acknowledgments

This research was funded by the BMW Group. Statements in this paper reflect the authors' views and do not necessarily reflect those of the funding body or former partners of the UDRIVE project.

6. References

[1] Vollrath, M., Huemer, A. K., Teller, C., et al.: 'Do German drivers use their smartphones safely?—Not really!', Accid. Anal. Prev., 2016, 96, pp. 29-38

[2] Prat, F., Planes, M., Gras, M. E., et al.: 'An observational study of driving distractions on urban roads in Spain', Accid. Anal. Prev., 2014, 74, pp. 8-16

[3] Young, K. L., Lenné, M. G.: 'Driver engagement in distracting activities and the strategies used to minimise risk', Safety Sci., 2010, 48, pp. 326-332

[4] Kidd, D. G., Tison, J., Chaudhary, et al.: 'The influence of roadway situation, other contextual factors, and driver characteristics on the prevalence of driver secondary behaviors', Transp. Res. Part F, 2016, 41, pp. 1-9

[5] Drews, F. A., Yazdani, H., Godfrey, C. N., et al.: 'Text Messaging During Simulated Driving', Hum. Fact., 2009, 51, (5), pp. 762–770

[6] Yannis, G., Laiou, A., Papantoniou, P., et al.: 'Impact of texting on young drivers' behavior and safety

on urban and rural roads through a simulation experiment', J. Safety Res., 2014, 49, pp. 25-31

[7] Caird, J. K., Johnston, K. A., Willness, C. R., et al.: 'A meta-analysis of the effects of texting on driving', Accid. Anal. Prev., 2014, 71, pp. 311-318

[8] Rudin-Brown, C. M., Young, K. L., Patten, C., et al.: 'Driver distraction in an unusual environment: Effects of text-messaging in tunnels', Accid. Anal. Prev., 2013, 50, pp. 122-129

[9] Simmons, S. M., Hicks, A., Caird, J. K.: Safety-critical event risk associated with cell phone tasks as measured in naturalistic driving studies: A systematic review and meta-analysis', Accid. Anal. Prev., 2016, 87, pp. 161-169

[10] Young, K. L., Salmon, P. M.: 'Examining the relationship between driver distraction and driving errors: A discussion of theory, studies and methods', Safety Sci., 2012, 50, pp. 165-174

[11] Beede, K. E., Kass, S. J.: 'Engrossed in conversation: The impact of cell phones on simulated driving performance', Accid. Anal. Prev., 2006, 38, pp. 415-421

[12] Ishida, T., Matsuura, T.: 'The effect of cellular phone use on driving performance', IATSS Res., 2001, 25, (2), pp. 6-14

[13] Platten, F., Milicic, N., Schwalm, M., et al.: 'Using an infotainment system while driving – A continuous analysis of behavior adaptations', Transp. Res. Part F, 2013, 21, pp. 103-112

[14] Schneidereit, T., Petzoldt, T., Keinath, A., et al.: 'Using SHRP 2 naturalistic driving data to assess drivers' speed choice while being engaged in different secondary tasks', J. Safety Res., 2017, 62, pp. 33-42

[15] Tivesten, E., Dozza, M.: 'Driving context influences drivers' decision to engage in visual–manual phone tasks: Evidence from a naturalistic driving study', J. Safety Res., 2015, 53, pp. 87-96

[16] Naujoks, F., Purucker, C., Neukum, A.: 'Secondary task engagement and vehicle automation–Comparing the effects of different automation levels in an on-road experiment', Transp. Res. Part F, 2016, 38, pp. 67-82

[17] Stutts, J., Feaganes, J., Reinfurt, D., et al.: 'Driver's exposure to distractions in their natural driving environment', Accid. Anal. Prev., 2005, 37, pp. 1093-1101

[18] Huisingh, C., Griffin, R., McGwin Jr, G.: 'The Prevalence of distraction among passenger vehicle drivers: A roadside observational approach', Traffic Inj. Prev., 2015, 16, (2), pp. 140-146

[19] Precht, L., Keinath, A., Krems, J. F.: 'Identifying the main factors contributing to driving errors and traffic violations – Results from naturalistic driving data', Transp. Res. Part F, 2017, 49, pp. 49-92.

[20] Bärgman, J., van Nes, N., Christoph, M., et al.: UDrive Deliverable D41.1: The UDrive dataset and key analysis results, 2017

[21] Castermans, J.: UDrive Deliverable D30.1: Overview of the Data Collection, 2017

[22] Alexander, L., Lopes, B., Ricchetti-Masterson, K., Yeatts, K.: 'Cross-sectional studies', https://sph.unc.edu/files/2015/07/nciph_ERIC8.pdf, accessed 24 April 2018

[23] Heinig, I., Jansen, R., Neumann, I., et al.: UDrive Deliverable D41.1: The UDrive dataset and key analysis results – Appendix B: Annotation Codebook, 2017

7. Appendix A

Table 7 Description of mobile phone task categories

Mobile phone task category	Description
Conversation hand-held	Driver is talking on a hand-held mobile phone or has the phone up to ear as if listening to a phone conversation
Conversation hands-free	Driver is talking or listening on a mobile phone using a hands-free device, such as a headset, in-vehicle integrated system, or hands-free speaker phone
Texting/ browsing	Driver is pressing buttons or a touch screen on the mobile phone to create and/ or send a text message or to browse in the internet or phone applications
Reading hand-held	Driver is looking at the screen of the mobile phone and clearly reading something, without a physical interaction
Reading hands-free	Driver looks to the cell phone regularly, without holding it and without a physical interaction
Holding	Driver is holding a mobile phone, but not manipulating it and not reading something
Other	Driver is interacting with a mobile phone in some other manner (e.g., taking pictures)

Note. According to the UDRIVE annotation codebook, see Heinig et al. [23].

A preliminary simulator study to investigate the effects of digital mirror failures on driving and glance behaviour, situation awareness, criticality and trust

Sanna M. Pampel [1*], Toby Nofal [1], Gary Burnett [1]

[1] Human Factors Research Group, University of Nottingham, University Park, Nottingham, UK
[*]sanna.pampel@nottingham.ac.uk

Abstract: Camera-based 'rear-view' displays within vehicles can improve aerodynamics and the field of view. However, digital technology may fail. Specifically for lane change situations, malfunctions may result in insufficient visual information and unsafe manoeuvres. Moreover, a degraded source may lead to distraction, compromised trust and thus lower acceptance. A driving simulator experiment aimed to determine the impact of a digital mirror failure on driving and visual behaviour, situation awareness (SA), criticality ratings and trust. Therefore, the existing 'wing mirrors' were replaced with in-vehicle LCD screens. In three drives in a UK motorway scenario, 19 drivers were instructed to perform ten lane-changes. During the second drive, the right (offside) digital mirror failed immediately after the instruction to move from the middle to the right ('fast') lane. Results show that the failure led to larger speed variation, more rear-view-mirror and slightly more over-the-shoulder checks, but increased observations of the right (failed) mirror, indicating distraction. Cumulative SA was not affected, but ratings for instability, complexity and variability increased. Drivers also recognised the heightened criticality. Unsurprisingly, trust decreased, potentially motivating the compensatory behaviours. In the third drive, which was free from failures, behaviours, criticality and trust returned to pre-failure levels, indicating no persistent long-term effects.

1. Introduction

The concept of mirrorless cars involves the replacement of traditional side mirrors with camera-based displays placed within vehicles, thereby improving vehicle aerodynamics and improving the field of view. Technological advancements mean that modern in-vehicle electronics are generally robust and highly reliable, with current systems able to successfully replace or augment aspects of vehicle control, such as braking and steering [1]. Nevertheless, digital technology may fail. A failure is defined as "an event that occurs when the delivered service deviates from correct service" [2, p. 2]. Hence, a failure constitutes the situation in which a system is not doing what it is intended to do. Besides faults related to the software and electronic circuits, camera-based systems are also susceptible to environmental factors that may limit the camera's vision, such as rain, dirt and ice, sun glare, or image distortions in low sunlight conditions. Despite the most diligent efforts to ensure the correct functioning of digital mirrors, designers need to envision scenarios in which a failure occurs. In the case of digital mirrors, it could potentially cause a frozen, blank or otherwise incorrectly displayed image. Specifically, for situations in which drivers' awareness of the sides and back of their car depends on digital mirrors, malfunctions (or excessive dirt / sun glare) may result in insufficient visual information and unsafe manoeuvres. Moreover, display failures may lead to significant levels of distraction, as drivers may (repeatedly) attempt to extract information from a degraded or even misleading source. In order to measure the impact of failures, Neukum and Krüger [3] developed a criticality scale, assessing the subjectively experienced degree of disturbance, ranging from imperceptible to uncontrollable, along with an 11-point scale, shown in Table 1.

Ultimately, negative experiences can compromise trust, which is "…the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" [4, p. 54]. Driving provides many such uncertain situations, in which drivers depend on mirror images to build sufficient awareness before making decisions. Decreased trust can then impact on the acceptance of technology [4-6]. For instance, numerous accounts of railway and aviation accidents resulting from the ignorance of alarms [cf. 7] illustrate how a lack of trust can lead to dangerous disuse. In addition, it is evident that trust is inversely related to the extent a device is monitored [6]. Hence, trust is particularly important for systems that provide a substitute for well-established, essential devices (such as a side mirror for a vehicle).

Table 1 Criticality rating scale

uncontrollable	dangerous			unpleasant			harmless			imperceptible
10	9	8	7	6	5	4	3	2	1	0

1.1. The current study

The current study aimed primarily to determine whether drivers responded to a digital mirror failure with compensatory behaviours and changes in self-reported trust. In terms of the former, they could change their speed and adjust their visual search such as using the rear-view mirror or conducting over-the-shoulder checks. Moreover, in order to better understand these effects, the research also aimed to investigate impacts of a failure on further subjective measures including situation awareness (SA) and criticality ratings. Because of the low likelihood of a digital mirror failure, repeated occurrences were not included in the present study.

Of interest to this analysis were the lane changes in which failures occurred, as well as the corresponding lane changes in the drives without failures. This was decided in order to measure effects of failures on subsequent mirror use when the mirror is functioning correctly.

2. Methodology

2.1. Mirror Failure Condition

In order to measure the effects of digital mirror failures, the drivers were subjected to a failure condition of the right (offside) digital mirror. The failure occurred at a dedicated but unpredictable time, immediately after being instructed to move from the middle into the right hand ('fast') lane, followed by subsequent lane change instructions. The failure always occurred during Drive 2 and involved the mirror turning blue for approximately 1 second followed by a frozen image with a road clear of traffic being presented, shown in Fig. 1.

Fig. 1. *Frozen image displayed in the right-hand mirror when the failure occurred*

2.2. Design

The study was conducted with a repeated-measures design, with one factor, Drive. This factor consists of three levels, Drive 1 to 3. The first Drive was a baseline Drive, where no failures occurred. During the second Drive, the failure occurred at a dedicated, but for the participant unpredictable time and remained until the end of this Drive. During the third Drive, no failures occurred, to measure whether the participants displayed any residual behaviours and attitudes that reflect carry-on effects after experiencing failure.

2.3. Apparatus

The experiment was conducted using a busy UK motorway scenario in a medium-fidelity driving simulator at the University of Nottingham. The simulator is normally equipped with external LCD wing mirrors, but for the current study these were replaced with separate LCD panels inside the vehicle, as shown in Fig. 2. The rear-view mirror remained unchanged. The right-hand screen was connected to an HDMI switch, so the experimenter was able to change the screen input. This meant the screen briefly flashed blue due to the temporarily missing signal, followed by an image emulating a frozen motorway scene, as shown above.

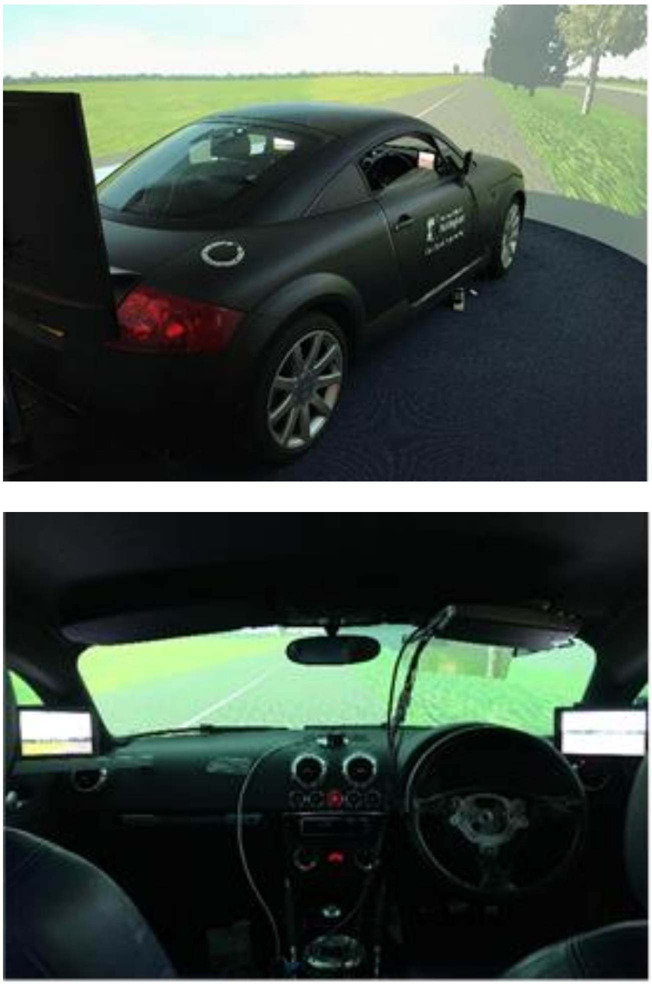

Fig. 2. *University of Nottingham driving simulator* **(a)** Fixed-base driving simulator **(b)** Digital mirror setup

2.4. Participants

Participants were recruited via an advertisement email to the staff and postgraduate students at the university as well as personally contacting colleagues and friends. In total, 19 regular drivers participated in the study, ranging from the age groups 18-29 to 60-69, and an average annual mileage of 3,516 miles (SD = 3,059 miles). As a gesture of appreciation, the participants were handed £10 shopping vouchers.

2.5. Procedure

At the beginning of the session, the participants were briefed on the study, without being informed about the failures, to avoid expectation. The drivers were then asked to fill in a consent form and a demographic questionnaire. The experiment involved three separate Drives (each approximately 10 minutes long). In each Drive, the participants were instructed to perform several lane-change manoeuvres while being surrounded by ambient traffic. These were delivered by voice instructions, which had been pre-recorded and were automatically played at specified distances down the road. The failure was triggered manually by the experimenter with a button press. Due to expected different speeds of the participants, it was not possible to closely control the location of the cars in the adjacent lane in relation to the participant vehicle. The lane change manoeuvres and the location of the mirror failure are illustrated in Fig. 3. Before the completion of the session, the participants were debriefed and it was explained to them that the purpose of the study involved the digital mirror failures.

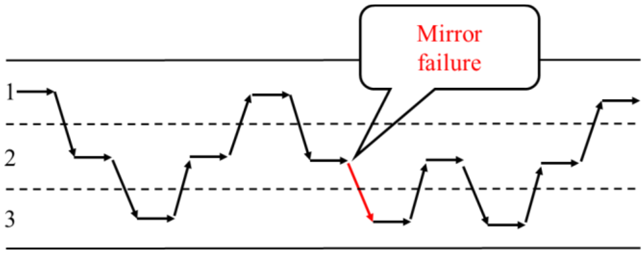

Fig. 3. *Plan view of the motorway with lane changes, showing the placement of the mirror failure within Drive 2*

2.6. Measures

Participants' reactions were recorded by the driving simulator software, operationalised as the speed and speed variation and lane position, as well as cameras inside the vehicle. The video recordings were then coded to identify glances into the digital mirrors, the rear-view mirror and over-the-shoulder checks. SA was measured with a 12-item questionnaire by Taylor and Selcon **[8]**. Trust was measured with a questionnaire by Jian et al. **[9]** and criticality with the criticality rating scale **[3]**.

2.7. Analysis

Of interest to this analysis was the lane change in which the failure occurred (Drive 2), as well as the corresponding lane changes in the Drives without failures (Drives 1 and 3). The time window for data gathering was from the onset of the failure until the successful completion of the lane change manoeuvre. If no lane change occurred, the data window lasted until the following lane change instruction.

The analysis was conducted with SPPS, using multivariate ANOVAs with Drive as within-subjects factor. In case the assumptions of parametric tests were violated, a Friedman test was performed instead, with Wilcoxon signed-rank tests for pairwise comparisons. All pairwise comparisons were Bonferroni-corrected.

3. Results

3.1. Driving measures

When the mirror right failed, six drivers did not perform the lane change that was instructed at that time. One of these drivers then also omitted the corresponding lane change in Drive 3. Generally, the drivers did not change their mean speed following the failure ($p = .150$). However, the analysis of the standard deviation of speed produced a main effect [$F(2, 36) = 3.45$, $p = .043$], which was due to larger speed changes in Drive 2 (mean = 10.22 m/s, SD = 4.08 m/s) compared to Drive 3 (mean = 7.16 m/s, SD = 3.01 m/s, $p = .025$). There was a main effect for the lateral variation [$F(1.234, 22.217) = 4.41$, $p = .040$], but post-hoc comparisons did not flag up significant differences.

3.2. Glance Behaviour

Only 4 of the 19 drivers performed a check over their shoulder in Drive 2, when the failure occurred, which was still more compared to 2 participants in Drives 1 and 3. However, due to the small numbers, this variable was not statistically analysed. Friedman tests of the mirror glances identified main effects for the number of glances to the right [$\chi2(2, N = 19) = 20.48$, $p < .001$] and rear mirrors [$\chi2(2, N = 19) = 21.26$, $p < .001$]. Pairwise comparisons showed an increase of glances into the right mirror by 113% from Drive 1 to 2 ($p = .003$), followed by a 51% decrease in Drive 3 ($p < .001$). Glances into the rear-view mirror increased by 184% from Drive 1 to 2 ($p < .001$) and then lowered by 63% in Drive 3 ($p = .003$). There were no significant pairwise differences between Drives 1 and 3.

3.3. Subjective SA

The cumulative SA score was higher on average in Drive 2 compared to the Drives without failure, but the effect was not significant ($p = .059$). When comparing the separate items, it was found that, from Drive 1 to 2, there were increases in instability ($p = .036$), complexity ($p =.024$) and variability ($p = .003$). Then, complexity decreased in Drive 3 ($p = .036$). No item produced a significant difference in SA between Drive 1 and 3.

3.4. Criticality

In Drive 1, the average critical rating was 2.79 and thus within the range of 'harmless'. An ANOVA of the criticality ratings produced a significant main effect [$F(1.21, 21.76) = 18.69$, $p < .001$]. Pairwise post-hoc comparisons assigned this effect to an increase in criticality ratings by 79% from Drive 1 to Drive 2 ($p = .004$) into 'unpleasant' as well as a subsequent decrease to 2.47 ('harmless', $p < .001$).

3.5. Subjective Trust

An ANOVA of the cumulative trust score resulted in a significant main effect [$F(2, 36) = 15.92$, $p < .001$]. Pairwise comparisons assigned this effect to a lowered trust score, by 67% from Drive 1 to 2 ($p < .001$), and a subsequent 141%

increase in Drive 3 (p = .001). Trust did not differ between Drive 1 and 3 (p = .359).

When considering the separate questionnaire items, pairwise Wilcoxon signed-rank tests showed that ratings worsened from Drive 1 to 2 for wariness (p = .021), harmfulness (p = .006), confidence (p = .003), dependability (p = .012), reliance (p = .001) and trust (p < .001). Answers then improved in Drive 3 for wariness (p = .033), harmfulness (p = .003), integrity (p = .042), reliance (p = .012) and trust (p = .001). Box plots of results are provided in Fig. 4.

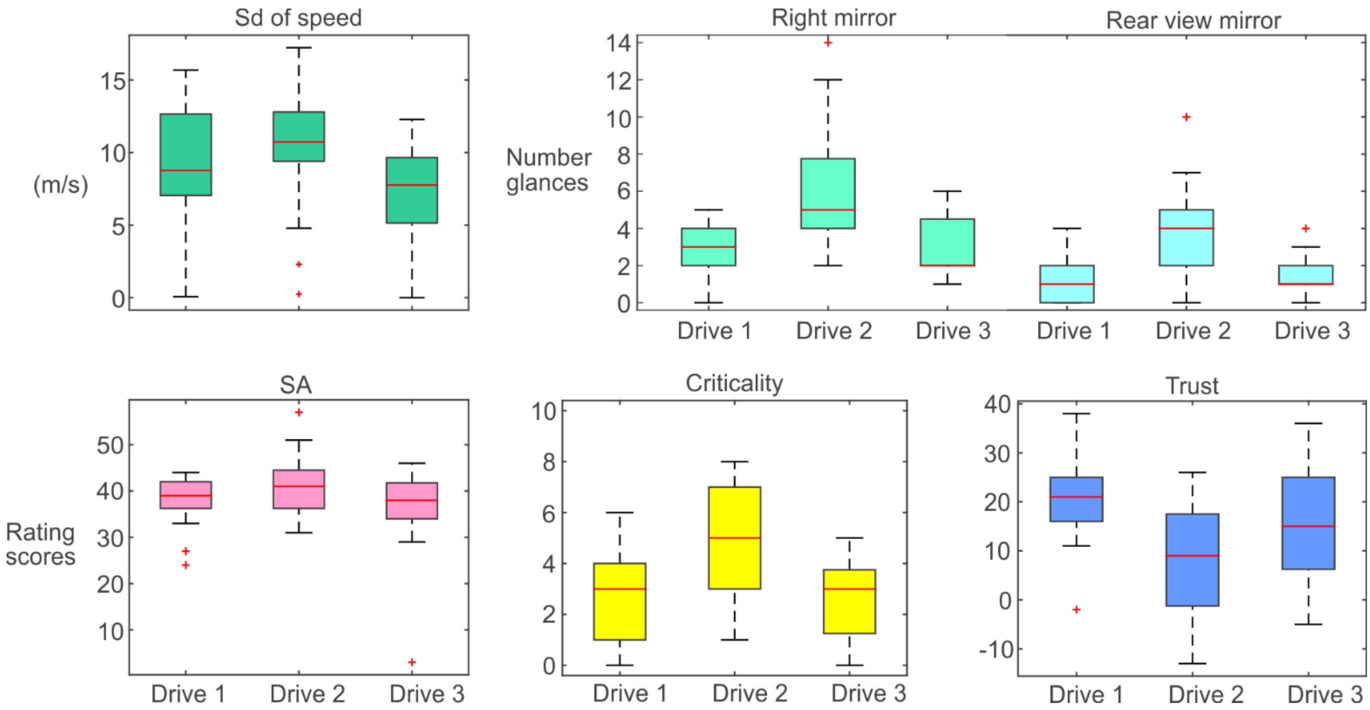

Fig. 4. *Box plots of summary measures*

4. Discussion

The present study investigated the effects of a ‘frozen-image’ failure of the digital mirror system on driving and visual behaviour, SA, criticality ratings and trust, measured in a driving simulator study supplemented with video recordings and questionnaires. Results show that the failure led to significant changes in behaviours. Although mean speed and lateral variation were not significantly affected, speed variation was higher following the failure (leading to non-significant decreases in mean speed). The drivers also compensated by looking more often into the rear-view mirror. Using the centre mirror seemed to have been the first course of action for the drivers, once they realised the failure. A slight increase in over-the-shoulder (blind-spot) checks could also be observed, but the number was generally unexpectedly low. It is a possibility that the driving simulator environment did not provide the visual experience that is realistic enough to support such checks, even during a mirror failure. However, an analysis of lane changes during a naturalistic driving study in the US [10] supports the observation that drivers tend to rely on rear-view-mirrors, more than on the respective side mirror, and the least on blind-spot checks. It has indeed been shown that brief rear-view-mirror checks decrease crash and near-crash risk [11]. Hence, possibly due to these compensatory behaviours, cumulative SA was not significantly affected, but the individual items: instability, complexity and variability were increased. It also appears that the drivers recognised the heightened criticality, rising from ‘harmless’ to ‘unpleasant’. The finding that the participants looked at the right (failed) mirror more indicates a potential distraction effect [12, 13]. The frozen image can be misleading, but the flashing blue screen preceding the frozen image might have mitigated that effect. The clarity of the situation was indicated by the timely increase in compensatory behaviours. In addition, when prompted by the experimenter at the end of the session, 17 of the 19 participants mentioned the failure, and none of them explicitly attributed it to the driving simulator equipment. Hence, it is suggested that a clear warning symbol, which immediately communicated the mirror’s state to the driver, could be useful in the case of such a failure. In this way, it could help the drivers build a correct mental model of the situation, which can result in potentially safer and more appropriate reactions [14, 15].

Ultimately, despite the difficulties of the situation, no collisions occurred, but the experimenter observed several ‘near-misses’, highlighting a potentially increased crash risk when failures occurred. The fact that six drivers refused to change into the fast lane with a failed mirror shows how these drivers prioritised safety, which is remarkable in the face of experimental instructions and the potentially associated social desirability [16].

The analysis of the trust questionnaire shows that trust in the digital mirrors was influenced by whether a failure occurred in a Drive or not, but only for the actual failure situation, not for the following failure-free Drive. In

Drive 2, trust in the technology decreased significantly, cumulatively and for most separate items. In summary, the mirror failure conditions significantly decreased self-reported trust. This adjustment in trust could have motivated the drivers to perform the compensatory behaviours, which were appropriate in this case.

There were no significant differences in any of the dependent variables between the first and third Drives, which were both free from failures. Hence, driving and visual behaviours, SA and perceived criticality returned to pre-failure levels when the digital mirror returned to normal functioning, but the reconstruction of previous trust levels is especially interesting. The finding that the impact on trust did not influence the later Drive can indicate that trust, in situations with a functioning mirror, is not influenced by earlier failures. However, the trust construct measured in questionnaires is considered potentially weak, and does not always translate into actual behaviour [17]. Another possible explanation for the restoration of trust involves an increased general exposure of the society to technology and therefore a higher level of initial trust [18]. In addition, even if people's expectations of a system are not met during the first uses, the expectations may be simply adjusted, so that trust is not necessarily affected [19].

5. Conclusions

The findings of the current study show how drivers may react when digital mirrors fail, particularly in critical situations such as lane changes. When a failure occurred in the simulator, the drivers performed compensatory behaviours such as changing their speed and performing more glances into rear-view mirrors, and thus maintained some degree of SA. However, the alternative mirror views do not provide sufficient information about the driver's side view of the car and the number of necessary over-the-shoulder checks was low. At the same time, increased glances into the failed mirror indicate its distracting effect. Subjectively, drivers rated the criticality of the situation as 'unpleasant' and indicated lowered trust in the technology. Behavioural and subjective measures, including trust, were restored once the mirror returned to full functionality, suggesting no lasting effects of the failure. Future research needs to investigate digital mirror failures in the real world, because a driving simulator study is only able to deliver initial indications, particularly as the graphics cannot replace a real-world view. A wider range of different manoeuvres can further aid the understanding of mirror use and responses to failures. It also needs to be considered whether a frozen image without an obviously flashing blue screen beforehand can be more difficult to realise and thus misleading and distracting. On the flipside, a permanent blue screen or clear failure symbols could mitigate distraction and motivate better compensatory actions.

6. Acknowledgments

The authors thank all the volunteers for participating in this study.

7. References

1. Nolte, T., H. Hansson, and L.L. Bello, *Automotive communications-past, current and future*, in *Emerging Technologies and Factory Automation, 2005. ETFA 2005. 10th IEEE Conference on*. 2005: Catania, Italy. p. 985-992.
2. Avizienis, A., J. Laprie, and B. Randell, *Fundamental Concepts of Dependability*. 2000, UCLA: Los Angeles, CA.
3. Neukum, A. and H.P. Krüger, *Fahrerreaktionen bei Lenksystemstörungen–Untersuchungsmethodik und Bewertungskriterien.* VDI-Berichte, 2003. **1791**: p. 297-318.
4. Lee, J.D. and K.A. See, *Trust in Automation: Designing for Appropriate Reliance.* Human Factors: The Journal of the Human Factors and Ergonomics Society, 2004. **46**(1): p. 50-80.
5. Dzindolet, M.T., S.A. Peterson, R.A. Pomranky, L.G. Pierce, and H.P. Beck, *The role of trust in automation reliance.* International Journal of Human-Computer Studies, 2003. **58**(6): p. 697-718.
6. Muir, B.M. and N. Moray, *Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation.* Ergonomics, 1996. **39**(3): p. 429-460.
7. Sorkin, R.D., *FORUM: Why are people turning off our alarms?* The Journal of the Acoustical Society of America, 1988. **84**(3): p. 1107-1108.
8. Taylor, R.M. and S.J. Selcon, *Subjective measurement of situational awareness*, in *Designing for Everyone, Proceeds of the 11th Cong International Ergonomics Association*, Y. Quéinnec and F. Daniellou, Editors. 1991, Taylor & Francis: London, UK.
9. Jian, J.-Y., A.M. Bisantz, and C.G. Drury, *Foundations for an Empirically Determined Scale of Trust in Automated Systems.* International Journal of Cognitive Ergonomics, 2000. **4**(1): p. 53-71.
10. Lee, S.E., E.C. Olsen, and W.W. Wierwille, *A comprehensive examination of naturalistic lane-changes*. 2004, National Highway Traffic Safety Administration: Washington, DC.
11. Klauer, S.G., T.A. Dingus, V.L. Neale, J.D. Sudweeks, and D.J. Ramsey, *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data*. 2006, National Highway Traffic Safety Administration: Washington, DC.
12. Fairclough, S., M. Ashby, and A.M. Parkes, *In-vehicle displays, visual workload and usability evaluation.* Vision in vehicles, 1993. **4**: p. 245-254.
13. Iqbal, S.T., X.S. Zheng, and B.P. Bailey, *Task-evoked pupillary response to mental workload in human-computer interaction*, in *CHI '04 Extended Abstracts on Human Factors in Computing Systems*. 2004, ACM: Vienna, Austria. p. 1477-1480.
14. Norman, D.A., *Some observations on mental models*, in *Mental models,* D. Gentner and A.L. Stevens, Editors. 1983, Lawrence Erlbaum Associates Inc.: Hillsdale, New Jersey, USA.

15. Norman, D.A., *The design of future things*. 2009, New York, NY: Basic books.
16. Zizzo, D.J., *Experimenter demand effects in economic experiments.* Experimental Economics, 2010. **13**(1): p. 75-98.
17. Ajzen, I., *Perceived Behavioral Control, Self-Efficacy, Locus of Control, and the Theory of Planned Behavior1*. Journal of Applied Social Psychology, 2002. **32**(4): p. 665-683.
18. Mcknight, D.H., M. Carter, J.B. Thatcher, and P.F. Clay, *Trust in a specific technology: An investigation of its components and measures*. ACM Transactions on Management Information Systems, 2011. **2**(2): p. 1-25.
19. Lankton, N., D.H. McKnight, and J.B. Thatcher, *Incorporating trust-in-technology into Expectation Disconfirmation Theory.* The Journal of Strategic Information Systems, 2014. **23**(2): p. 128-145.

Efficient monocular point-of-gaze estimation on multiple screens and 3D face tracking for driver behaviour analysis

Jon Goenetxea *, Luis Unzueta, Unai Elordi, Juan Diego Ortega, Oihana Otaegui

Intelligent Transport Systems and Engineering Department, Vicomtech, Paseo Mikeletegi 57, Donostia, Spain
*jgoenetxea@vicomtech.org

Abstract: In this work, we present an efficient monocular method to estimate the point of gaze (PoG) and the face in the 3D space of multi-screen driving simulator users, for driver behaviour analysis. It consists in a hybrid procedure that combines appearance and model-based computer vision techniques to extract the 3D geometric representations of the user's face and gaze directions. These are placed in the same virtual 3D space as those of the monocular camera and the screens. In this context, the intersections of the overall 3D gaze vector with the planes that contain each screen is calculated with an efficient line-plane intersection geometric procedure. Finally, a point-in-polygon strategy is applied to see if any of the calculated PoGs lies within any of the screens, and if not, the PoG on the same plane as that of the closest screen is provided. Experiments show that the error for the obtained PoG accuracy is reasonable for automotive applications, even in the uncalibrated case, compared to other state-of-the-art approaches, which require the user's calibration. Another advantage is that it can be integrated in devices with low computational capabilities, such as smartphones, with sufficient robustness for driver behaviour analysis.

1. Introduction

Typically, state-of-the-art eye gaze estimation techniques obtain the point of gaze (PoG) on one screen, only [1]. However, in the case of driving simulators there are usually more than one, e.g., one for the front view, one for each side view, another one for the dashboard, etc (Fig. 1). Besides, there can be different objects of interest at different locations of each screen and obtaining the gaze fixations and saccades, derived from the PoG, accurately on each screen is important for driver behaviour analysis [2]. Additionally, it is also preferable to simplify the installation and calibration of sensors and to reduce the power consumption as much as possible, avoiding alternative possibilities such as placing a dedicated PoG estimator for each screen. Thus, in our context, we only consider one monocular camera in front of the user and a humble CPU, e.g., those included in an embedded PC or a smartphone.

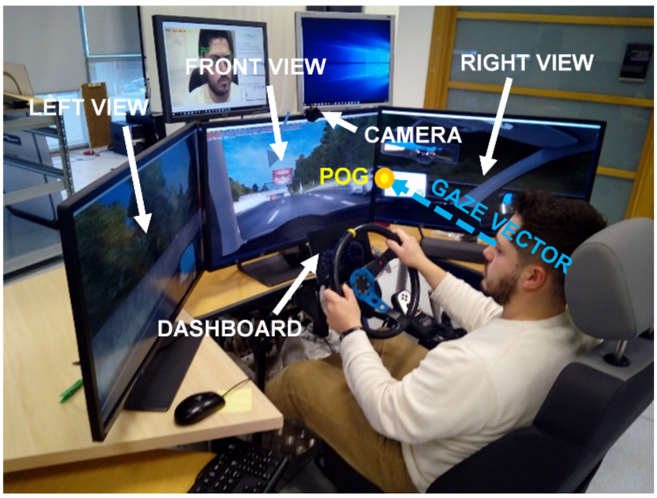

Fig. 1. *Multi-screen simulator setup for driver behaviour analysis, based on human-machine interaction, including PoG and 3D face tracking*

In automotive platforms, visual features of the face and eye regions of a driver provide cues about their degree of alertness, perception and vehicle control. Knowledge about driver cognitive state helps to predict, for example, if the driver intends to change lanes or is aware about obstacles and thereby avoid fatal accidents. These systems use eye tracking setups mounted on a car's dashboard along with computing hardware running machine vision algorithms, with computational capabilities far below from those of off-the-shelf desktop PCs. Major sources of error in automotive systems arise principally from platform and user head movements, variable illumination, and occlusion due to shadows or users wearing glasses, which need to be handled robustly but also efficiently due to the computational constraints.

The current state of the art of eye gaze estimation systems applied to automotive platforms includes different kind of approaches and uses. There are approaches that consider eye movement features (e.g., fixations, saccades, smooth pursuits, etc) for deriving driver cognitive states, such as driver distraction [3]. Other approaches apply classification techniques to eye images related with different gaze zones, to detect where the driver is looking at while driving [4]. There are also approaches that track facial features, 3D head poses and gaze directions relative to the car geometry to detect eyes-of-the road condition of the driver [5]. Other approaches study the driver's gaze behaviour (e.g., glance frequency and glance time) to evaluate the driving performance when they interact with other devices (e.g., a portable navigation system) while driving [6]. Finally, there are also approaches that study the dynamics between head pose and gaze behaviour of drivers to predict gaze locations from the position and orientation of a driver's head [7] or to categorise different kind of driver behaviours while driving [8].

Our main motivation in this work is to increase the grade of sophistication of all this kind of use cases by developing a more accurate, more robust, but still efficient method for estimating the head pose and eye gaze of drivers, compared to previous approaches. We paid special attention to the case of multi-screen simulators, where the relation between the PoG and the rendered graphics can be directly established, and therefore, richer data could be extracted for behaviour analysis. In order to do so, it is necessary to relate the 2D image projections of the driver's facial and ocular cues, captured from the monocular camera, with the 3D space. Ideally, this would require not only obtaining the person's 3D eye gaze vectors from the images, but also the person's 3D eye positions and the surrounding potential targets' geometries in the same 3D space, the camera characteristics from which that space is observed, and an additional calibration stage done by the user. However, in many applications it is not easy to obtain all these data. This is the case of automotive applications, where it is not comfortable for the driver to spend time calibrating the eye gaze system. Other important factors are that the estimated gaze vector should have a low level of noise, but it should still be sensitive to quick eye movements, and that the estimated gaze vector should be robust to head movements, which in the case of driving, normally happen many times.

Our approach to tackle all these factors consists in a hybrid procedure that combines appearance and model-based computer vision techniques to extract the 3D geometric representations of the user's face and gaze directions. These are then placed in the same virtual 3D space as those of the monocular camera and the screens. This reconstructed virtual 3D world is where the driver's behaviour can then be analysed, based on the estimated PoG on the different targets of the scene and the 3D head pose, without necessarily requiring calibration data. It has been designed to have an acceptable balance between accuracy, robustness and efficiency, so that it can be integrated into devices with low computational capabilities that might be used in vehicles.

The rest of the paper is organised as follows. Section 2 introduces the proposed hybrid system. Section 3 illustrates details about our experiments and presents some discussions about them. Section 4 concludes the paper.

2. Methodology

The methods to estimate the eye gaze from monocular images and videos can be categorised in two types of approaches: model-based [5][9] and appearance based [4][8][10][11][12][13][14]. Next, we study more in detail the pros and cons of each and then we explain our proposed hybrid approach.

2.1. Model-based vs appearance-based

The model-based approach relies explicitly in 3D graphical models that represent the geometry of the eye (typically as spheres) which are fitted to the person's detected eye features in the image (typically, the iris and the eye corners). Thus, the fitted 3D model allows inferring the 3D eye gaze vector, which is then used to deduce where the person is looking at. These methods imply some drawbacks, such as: They require to precisely locate the iris of the eye in the image; this is often impossible, for example when the user's eyes are not wide open, which is the normal case. In order to estimate the eye gaze direction, they need the user's head coordinates system as reference. Therefore, the success of these methods is highly dependent on the precision with which the user's head coordinates system has been localised. Besides, although simple, they require an initialisation scheme: the user needs to intentionally look at one or more points on a screen. Otherwise, eye vectors cannot be obtained with sufficient precision. In sum, since they are pure geometric methods, their precision is strongly dependent on the precision of the estimated eyeball and pupil centres. However, common images do not enable to obtain this information with high precision.

On the contrary, the appearance-based approach establishes a direct relation between the person's eye appearance and the corresponding eye gaze data of interest (e.g., the 3D eye gaze vector) by applying machine learning techniques. Thus, a dataset of annotated images is used to train a regression model, which is then used to deduce where the person is looking at, when applied to the person's eye image extracted from the image.

In the last few years, the appearance-based methods have been greatly benefited by the revolutionary results obtained by the emerging deep learning techniques in computer vision applications and have become the current state of the art in the field. They allow to generalise much better the learned relation between the eye appearance and the corresponding eye gaze data than alternative machine learning approaches (based on "handcrafted" image features and "shallow" layered learning architectures), when a huge dataset of annotated images is used for training. Typically, hundreds of thousands or even millions of samples are used, which may include real data [10][14], photorealistic synthetic data [11][13] or even a mixture of both [12]. This way, eye gaze direction estimation systems can obtain better accuracies with people whose appearance has not been included in the training of the regression model.

However, an effective eye gaze direction estimation system does not only require obtaining accurate eye gaze data from eye images, but it also requires applying properly the eye gaze data to the environment, so that it is possible to deduce where the person is looking at.

2.2. Hybrid approach

Fig. 2 shows the general overview of the workflow of our approach, where the inputs are a monocular image grabbed by one camera in front of the user, a parametric deformable 3D face model (Fig. 3), the camera intrinsic parameters and the screen geometries. The outputs are his/her estimated PoG with respect to the considered screens and his/her facial mesh in the 3D space, which includes information about his/her head position, orientation and expression. In this workflow, we distinguish three blocks: (1) the 3D face model adjustment to the user's face image, (2) the normalisation of the 3D gaze estimation and (3) the estimation of the eye gaze direction with respect to the targets.

The first block comprises computer vision procedures to detect and track facial regions on the image, localise facial landmarks and fit the 3D face model to those landmarks, by optimising the following objective function:

$$e = \arg\min \frac{1}{n}\sum_{j=1}^{n}[d_j - p(f, w, h, \boldsymbol{t}, \boldsymbol{r}, \boldsymbol{s}, \boldsymbol{a})_j\ ]^2 \qquad (1)$$

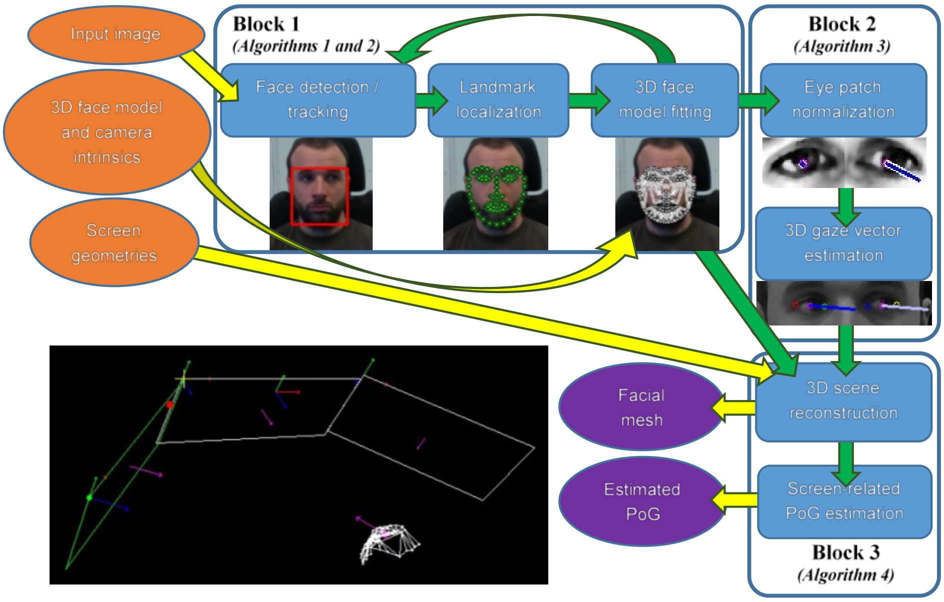

Fig. 2. *Workflow of the multi-planar PoG estimation and 3D face tracking approach*

where:

- $\boldsymbol{d}$ = {d_1, d_2, d_3, …} are the detected 2D landmark positions.
- $\boldsymbol{p}$= {p_1, p_2, p_3, …} are the 2D projections of the corresponding 3D deformable model vertices. $\boldsymbol{p}$ is a function that depends on the camera parameters (f, w, h) and on the parameters of the graphical object ($\boldsymbol{t}$, $\boldsymbol{r}$, $\boldsymbol{s}$, $\boldsymbol{a}$). Function $\boldsymbol{p}$ represents the 2D projections on a surface of vertices, which are 3D. The goal is to minimise the distance between the detected 2D landmark positions in the image and the vertices of the projections.
- f is the focal length of the camera from which the image was obtained.
- w is the image pixel width.
- h is the image pixel height.
- $\boldsymbol{t}$= {t_x, t_y, t_z} are the XYZ positions of the face model with respect to the camera.
- $\boldsymbol{r}$= {r_x, r_y, r_z} are the roll-pitch-yaw rotation angles of the face model with respect to the camera.
- $\boldsymbol{s}$= {s_1, s_2, s_3, …} are the shape-related deformation parameters.
- $\boldsymbol{a}$= {a_1, a_2, a_3, …} are the action-related deformation parameters.
- n is the number of 2D landmark positions.
- e is the residual error.

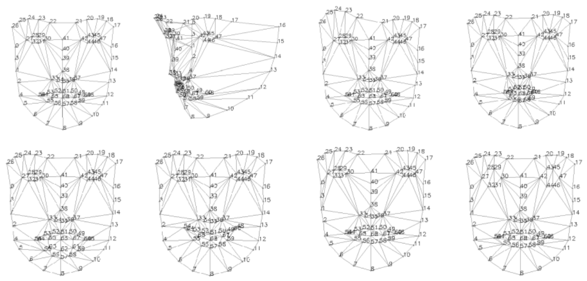

Fig. 3. *A generic deformable 3D face model and some of its deformation parameters compatible with our method*

For the localisation of the user's face region two stages are distinguished: (1) the initial face detection and posterior re-detections when the tracking is lost, and (2) the in-between face tracking. This is relevant as tracking algorithms typically are more efficient and require less memory than those for face detection. Thus, the face detection algorithm is only activated when the user's face is not being tracked. The detection is done with the SSD deep neural network [15], which has shown to be robust under challenging conditions, trained specifically with multiple-pose faces. The tracking is based on CLNF [16], applied at landmark level, which has a good balance between computational cost and localisation reliability and stability. The landmark distribution is constrained by a parametric 3D face model, to avoid impossible human facial shapes. The tracking is considered to be lost when the image under the face region does not correspond to a human face, according to the learned face pattern (see Algorithms 1 and 2 for further details).

Algorithm 1: Hybrid face model detection-tracking fitting algorithm

Input: ***The image sequence I***
Output: ***The face model parameters {t, r, s, a} that overlap the model to the user's face, throughout I***

```
For each I_j ∈ I do
  if Face detection needed then
    Reset the face model parameters of the graphical model to the neutral configuration
    Run the face region detector in the image
    Store the detected user's face image patch and face region
  else
    Locate a stored face image patch in the image (via pattern matching)
    Verify that the located patch corresponds to a real face (via pattern classification)
    if Located face region contains a real face then
      Store the located face region
    end
  end
  if Face region available then
    Run the face landmark detector in the face region
    Adjust the 3D face model to the detected landmarks (Algorithm 2) → {t, r, s, a}_j
  end
end
(Optional) Filter {t, r, s, a} with an appropriate approach for face movements
```

Algorithm 2: Three-stage face model adjustment algorithm

Input:
- ***Set of 2D landmark positions d in the image***
- ***The relation list between the landmark and vertices***
- ***The camera parameters {f, w, h}***

Output: ***The face model parameters {t, r, s, a} that overlap the model to the user's face***

Set the deformation parameters $\{\boldsymbol{s}, \boldsymbol{a}\}$ to zero
Convert the current parameter values to the normalised range workspace
Optimise, using for example the Levenberg-Marquardt algorithm [17][18], Eq. (1) with $\{\boldsymbol{t}, \boldsymbol{r}\}$ as the only variables
Optimise, using for example the BFGS algorithm [19][20][21][22], Eq. (1) with $\{\boldsymbol{s}\}$ as the only variables
For each $a_k \in \boldsymbol{a}$ **do**
    Optimise, using for example the BFGS algorithm, Eq. (1) with $\{a_k\}$ as the only variable
end

Once the different facial parts are localised, the image regions around both eyes are extracted, and their shape and intensity distributions are normalised, so that a deep neural network, based on [10], can infer the corresponding 3D gaze vectors. Then, an overall gaze vector of the user is calculated as the weighted mean vector of both eyes with its origin at the midpoint of both eyes (see Algorithm 3).

Algorithm 3: Normalised left and right eye gaze vectors estimation algorithm

Input:
- ***The image sequence I***
- ***2D left $\{e_1, e_2\}_l$ and right $\{e_1, e_2\}_r$ eye corner landmark positions, throughout I***
- ***The adjusted face model geometry and parameters, throughout I***
- ***The pre-trained deep neural network for regressing 3D gazes from normalised eye images***

Output: ***The user's normalised left and right eye gaze vectors estimation $\{g_l, g_r\}_{norm}$, throughout I***

For each $I_j \in \boldsymbol{I}$ **do**
    Calculate M for each eye (Eq. (2))
    Obtain I_{norm}^{shape} for each eye (Eq. (3))
    Obtain I_{norm} for each eye (via image equalisation)
    Mirror I_{norm} for the eye not corresponding to that considered by the regressor (left or right)
    Process both I_{norm} with the pre-trained deep neural network
    Un-mirror the response for the mirrored eye image → $(\{\boldsymbol{g}_l, \boldsymbol{g}_r\}_{norm}^{reg})_j$
    Apply the dominant eye and head rotation's correction factor (Eq. (4)) → $(\{\boldsymbol{g}_l, \boldsymbol{g}_r\}_{norm}^{corrected})_j$
    Divide both regression results by their corresponding Euclidean norms → $(\{\boldsymbol{g}_l, \boldsymbol{g}_r\}_{norm})_j$
end

The affine transformation matrix M is calculated as follows:

$$\begin{bmatrix} \alpha & \beta & (1-\alpha)\cdot c_x - \beta \cdot c_y \\ -\beta & \alpha & \beta \cdot c_x + (1-\alpha)\cdot c_y \end{bmatrix} \qquad (2)$$

where:
- $\alpha = s \cdot \cos(\theta)$
- $\beta = s \cdot \sin(\theta)$
- $s = (w - 2 \cdot m_1)$
- θ refers the horizontal rotation angle of the line that connects both eye corners.
- $\{c_x, c_y\}$ are the image coordinates of the centre of rotation in the source image.

Then, the source image I_{input} is transformed, that is to say, normalised in shape, using the matrix M, as follows.

$$I_{norm}^{shape}(x, y) = I_{input}(M_{11}x + M_{12}y + M_{13}, M_{21}x + M_{22}y + M_{23}) \qquad (3)$$

It must be noted that the applied eye shape normalisation procedure usually results in distorted images; normally, the further the user's face is with respect to frontal viewpoints, i.e., the most distant eye's appearance may look, normally, the more distorted the images become.

As a matter of example, Fig. 4 shows three examples of the distortion that happens in the normalised appearance of distant eyes in non-frontal faces, when the head's yaw angle is changed. As can be observed, the green points do not match exactly the white ones because the deformability of the graphical object is not perfect. At most, e is minimised (Eq. (1)). Consequently, this distortion may affect in stability of the estimated gaze for different yaw rotation angles of the head. A similar instability may also happen for different pitch angles, but in a lower degree.

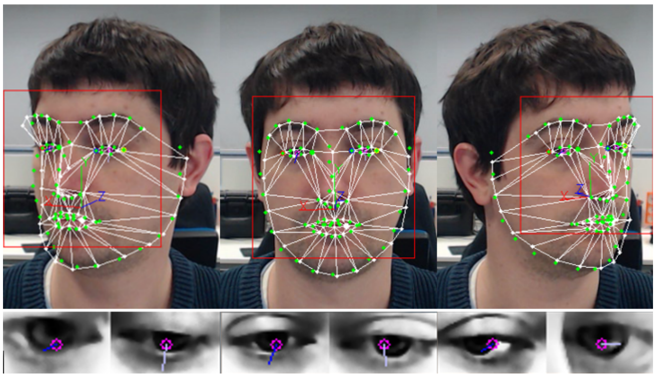

Fig. 4. *Examples of the distortion that happens in the normalised appearance of the most distant eyes in non-frontal faces, when the head's yaw angle is changed*

Thus, in order to reduce this effect, the vectors obtained in the previous step ($\{\boldsymbol{g}_l, \boldsymbol{g}_r\}_{norm}^{reg}$) are corrected by a factor that gives more importance to the dominant eye (the less distorted eye) and which is proportional to the head's pitch and yaw rotation angles, as follows:

$$\{\boldsymbol{g}_l, \boldsymbol{g}_r\}_{norm}^{corrected} = \{w_d \cdot \boldsymbol{g}_l, (1 - w_d)\cdot \boldsymbol{g}_r\}_{norm}^{reg} + \begin{Bmatrix} K_y \cdot (r_y - r_{y0}) \\ K_x \cdot (r_x - r_{x0}) \\ 0 \end{Bmatrix} \qquad (4)$$

where:
- w_d is the weight of eye dominance.
- r_{x0} is the reference pitch angle.
- r_{y0} is the reference yaw angle.
- K_x is the proportionality constant for the pitch angle.
- K_y is the proportionality constant for the yaw angle.

In the case of big out-of-plane head rotations where both eye images are too distorted to be reliable, the gaze estimation relies solely on the head direction. The values of these parameters and ranges are experimentally determined, depending on the final application. For instance, the reference pitch and yaw angles could be the average values from those observed during the image sequence, while the user's head poses are closer to frontal viewpoints, while the proportionality constants could be determined based on the observations of the gaze stability while the user is moving the head, but maintaining the point of gaze. Finally, each vector is divided by the Euclidean norm, so that to assure that the resulting vectors have unit norm, and this way both normalised gaze vectors are obtained.

It is remarkable that these 3D eye gaze vectors have been obtained without any previous calibration e.g. without any initialisation procedures. This is especially important in applications requiring real-time monitoring of the eye gaze, such as automotive applications.

Algorithm 4 shows how the eye gaze direction is estimated with respect to the targets. First, the target geometries are placed with respect to the camera's coordinate system, which is the same reference used for the face and eye gaze vectors, already estimated in previous blocks. The camera's coordinate system has been previously pre-established. In other words, it is assumed that the camera's coordinate system is well-known. A target is modelled or referred to as a set of polygons formed by k points $\boldsymbol{b}$ and lines $\boldsymbol{l}$, and their corresponding planar surfaces $\{\boldsymbol{v}, q\}$ (where $\boldsymbol{v}$ is the normal vector and q the distance from the origin) that define the objects that need to be related with the user's point of gaze (e.g., a screen is represented by a rectangular plane). Then, the 3D face model is placed in the scene with the obtained parameters. Then, the normalised left and right eye 3D gaze vectors are transformed, so that they are referred to the coordinate system of the camera (i.e., not to the normalised camera viewpoint, as before). This is done by removing the effect of the rotation angle θ that was used for the affine transformation applied to each normalised eye shape, like this:

$$\{g_l, g_r\} = \begin{Bmatrix} -\cos(\theta)\cdot(\{g_l, g_r\}_{norm})_x + \sin(\theta)\cdot(\{g_l, g_r\}_{norm})_y \\ -\sin(\theta)\cdot(\{g_l, g_r\}_{norm})_x - \cos(\theta)\cdot(\{g_l, g_r\}_{norm})_y \\ (\{g_l, g_r\}_{norm})_z \end{Bmatrix} \quad (5)$$

Then, both gaze vectors are combined by calculating its geometric mean g, which it is assumed to be the user's overall gaze vector. The gaze vector may optionally be filtered by taking into account its frame-to-frame motion and an appropriate filtering method for eye movements. The origin of this vector is preferably placed in the middle position (mean value) of both eye centres from the 3D face, Ɛ. Thus, the point of gaze PoG for each target plane can be estimated, like this:

$$\boldsymbol{PoG_t} = \mathcal{E} + \frac{(q - v\cdot\mathcal{E})}{v\cdot g}\cdot \boldsymbol{g} \quad (6)$$

Finally, a point-in-polygon strategy [23] is applied to see if any of the calculated PoGs lies within any of the screens. As can be observed, the point-in-polygon strategy may result in that the PoG goes through a polygon, or that it does not go through any polygon. If it does not go through a polygon, the method provides the closest polygon. For example, in line 11 of Alg. 4, if the PoG does not go through a polygon, the distance to the polygon is stored. And in line 12, the current measured distance is compared to the minimum measured distance (which is the stored one), in order to guarantee that the closest polygon is finally selected.

Algorithm 4: Target-related point of gaze estimation algorithm

Input:
- ***The set of polygons formed by k points b and lines l, plane normal vectors v and plane distances q with respect to the camera that represent the target objects*** $\{\boldsymbol{b_k}, \boldsymbol{l_k}, \{\boldsymbol{v}, \boldsymbol{q}\}\}_t$
- ***The adjusted face model geometry and parameters {t, r, s, a}, throughout I***
- ***The user's normalised left and right eye gaze vectors estimation*** $\{g_l, g_r\}_{norm}$***, throughout I***

Output: ***The user's PoG with respect to the targets in the scene, throughout I***

```
1:  Place the target polygons with {b_k, l_k, {v, q}}_t
2:  For each l_j ∈ I do
3:    Place the 3D face model with {t, r, s, a}_j
4:    Transform {g_l, g_r}_norm with Eq. (5) → {g_l, g_r}_j
5:    Calculate the geometric mean vector → g_j
6:    (Optional) Filter g_j with an appropriate approach for gaze
      movements
7:    d_t^min = BIG_NUMBER
8:    For each target t do
9:      Calculate the point of gaze PoG_t in the target's plane (Eq. (6))
10:     Apply a point-in-polygon strategy to the target's polygon
11:     If point-in-polygon test successful then → PoG_j and break
12:     Else store PoG_t and distance to polygon d_t
13:     If d_t < d_t^min then → PoG_j = PoG_t
14:       d_t^min = d_t → PoG_j = PoG_t
15:     end
16:   end
17: end
```

3. Results

We have evaluated our approach with an experiment where 8 people have been recorded by a camera in front of them, while using a driving simulator with three screens (Fig. 1). The participants were requested to look at different control points located at zones of interest on the screens: (1) left window, (2) left side mirror, (3) horizon, (4) road, (5) navigation panel, (6) rear mirror and (7) right side mirror (Fig. 5). They were free to rotate their head as they considered (no instructions were given about this). The accuracy of our approach has been measured in this setup without including a user-calibration stage. Thus, if the PoG obtained directly as explained above lies within the targeted zone of interest, it is

Fig. 5. *The considered zones of interest in the simulator to analyse the driver's PoG.*

Table 1 Comparison among different state-of-the-art eye gaze estimation systems and ours

Method category	Paper reference	Setup	Accuracy metrics (Mean % and/or º)
Model-based	[5]	1 camera, 1 IR illuminator and 18 gaze zones, considering day (no-glasses/glasses/sun-glasses) and night (no-glasses/glasses) scenarios	>95% on-the-road >90% off-the-road (for all scenarios)
	[9]	3 cameras (2 facing driver, 1 looking out) and 6 gaze zones	94.9%
Appearance-based	[4]	1 camera and 8 gaze zones	92.75%
	[8]	1 camera and 6 gaze zones	94.6%
	[14]	1 camera and 20 on-screen positions	10.8º (cross-dataset evaluation)
Hybrid	Ours	1 camera and 7 gaze zones	97.0% / 4.6º (front screen) 87.7% / 11.5º (side screens)

considered a correct response, wrong otherwise. Besides, we measured the angle between the vector that goes from the head to the targeted control point and from the head to the estimated PoG.

Table 1 shows the obtained results, along with those obtained by other state-of-the-art model-based [5][9] and appearance-based [4][8][14] alternatives with similar setups and conditions. Ideally, we would have reimplemented and adapted to our setup all these approaches so that then we could measure the differences under the same working conditions. However, taking into account that there are many implementation details that are not available in the publications, which can be important for the reproduction of the reported results, we have preferred to include them here directly with their corresponding setups and accuracy metrics. In some cases, they are given in degrees between the estimated and ground-truth gaze vectors and in other with a percentage of the number of times in which the correct gaze zones are reached. In our case, we provide both metrics so that it is easier to compare with the other approaches, despite the differences among the setups and conditions. Nevertheless, note that due to that reason this comparison is more qualitative than quantitative, except for their own setups with respect to their corresponding ground truth measurements.

[9] has the most different setup as it uses two cameras to capture the driver's data, and hence, it has the possibility of estimating 3D features directly and thus improve the accuracy, compared to the monocular case. However, we prefer to avoid this kind of setups in order to simplify the installation and configuration (i.e., calibration) and reduce the power consumption.

In the case of [5], it follows a similar scope to ours, using facial feature tracking, 3D head pose and gaze estimation, but with some relevant differences. The head pose estimation algorithm is based on the 'weak-perspective' assumption, which with the kind of images obtained in this setup produces an inherent error due the orthographic projection that needs to be compensated. On the other hand, its proposed gaze vector estimation procedure is model-based, which has the drawbacks already stated before.

Both [4] and [8] rely on classifiers trained with the relations between gaze zones and feature descriptors composed by 2D facial part and ocular image cues. The drawback of this kind of approaches is that, as they do not estimate 3D data, they need to be specifically trained for each setup, and provide more limited information for behaviour analysis.

[14] relies on a deep neural network to estimate the 3D gaze vector, in similar way as we do, but including both the normalized eye appearance and the head orientation as input data for the network. In this case, the approach is evaluated with people looking at a laptop screen, so no profile views are contemplated like those that occur in our case when users look at the side screens and the eye appearances get distorted.

In our case, it can be seen that we obtain sufficient accuracy to relate rendered graphics with the user's observations, despite not having calibrated the system for each user. As expected, the accuracy is lower for the side screens, but still high enough (Fig. 6). Anyway, these errors should be considered when designing the recognition areas for the interaction with the elements of the scene, i.e., for higher errors the area of interaction around the element should be bigger too. Fig. 7 shows that our approach can handle quick eye movements, but maintaining a low level of noise for fixations.

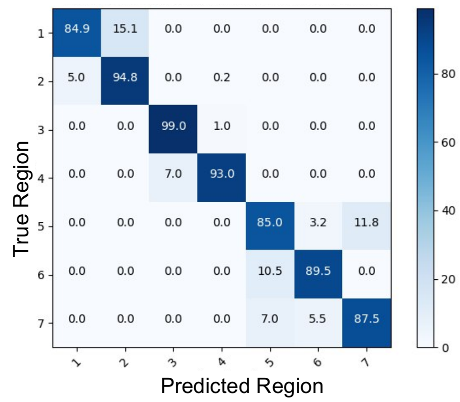

Fig. 6. *Confusion matrix of the predictions obtained by our approach for the considered gaze zones*

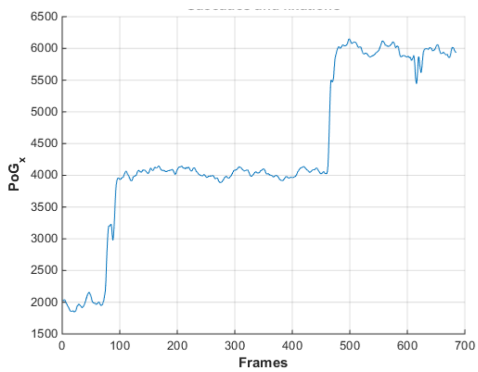

Fig. 7. *An example of PoG_x signal where saccades and fixations can be appreciated, along with the level of noise*

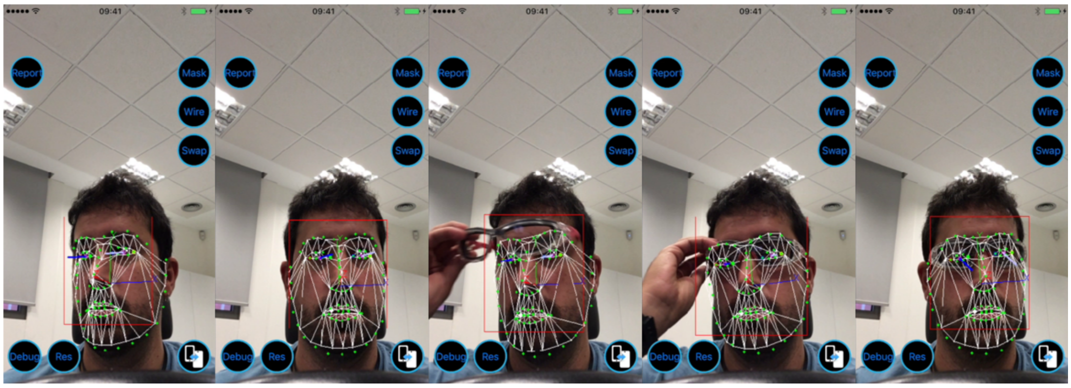

Fig. 8. Examples of the approach running in an iPhone SE, while the user puts thick glasses on and the system keeps working

On the other hand, in order to evaluate the efficiency of our approach and its suitability for running it in devices which can then be used in real vehicles, where one cannot expect installing CPUs/GPUs like those of desktop/laptop PCs, we have integrated it in an app for smartphones with iOS and Android operating systems (Fig. 8). It is remarkable to state that the operating system can also have an impact in the overall performance of the app, due to the multi-level structure and different programming languages in which the app needs to be programmed (i.e., the core of the approach is programmed in C++ for both operating systems, while the interface is in Objective-C for iOS and Java for Android). More specifically, we have tested the iOS app in an iPhone SE (with iOS 10.3.2) and the Android app in a Docomo smartphone (with Android 6). The measured average FPS (frames-per-second) of our app in each case has been 30 and 20 respectively, which reveals the efficiency and suitability of our approach to be applied in a real-world scenario.

4. Conclusion

One of the advantages of our approach is that with a simple setup we can efficiently estimate the PoG of the user in multiple screens of a simulator, allowing to relate directly the rendered graphics that represent the different elements of the scene with the user's observations.

Moreover, as the rendered scene simulates a physical car environment with a distribution close to a real case, this approach is suitable to be used inside a driving situation. In a real scenario, the zones in the rendered scene fit with the key attention zones considered while driving (with some variations depending on the car). This way it is easier to generate richer data for developing driving behaviour analysis approaches.

Another advantage is that it can be integrated, processed and executed in devices with low computational capabilities, such as smartphones.

Future work will principally focus on optimizing the deep neural network designs for the face detection, landmark localization and eye gaze vector estimation stages to further improve their efficiency in ARM-based CPUs.

We also plan to adapt the approach to real vehicle setups. The primary challenge in a real driving scenario is the illumination variability. Some image capture setups reduce the illumination issues using specific hardware like infrared cameras or a sort of optical filters. These changes call for particular datasets to re-train some of the models (e.g. eye gaze estimation model and face detection model), but the method pipeline is not affected.

5. Acknowledgments

This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 690772, VI-DAS project).

6. References

[1] Kar, A. and Corcoran, P.: 'A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms', IEEE Access, 5, 2017, pp. 16495–16519

[2] Kasneci, E., Kasneci, G., Kübler, T. C., et al.: 'Online recognition of fixations, saccades, and smooth pursuits for automated analysis of traffic hazard perception', Artificial Neural Networks, Springer Series in Bio-/Neuroinformatics, 4, 2015, pp. 411–434

[3] Liang, Y., Reyes, M. L., Lee, J. D.: 'Real-time detection of driver cognitive distraction using support vector machines', IEEE Trans. Intell. Transp. Syst., 8, (2), 2007, pp. 340–350

[4] Chuang, M.-C., Bala, R., Bernal, E., et al.: 'Estimating gaze direction of vehicle drivers using a smartphone camera', Proc. IEEE Conf. Comput. Vis. Pattern Recogn. Work. (CVPRW), 2014, pp. 165–170

[5] Vicente, F., Huang, Z., Xiong, X., et al.: 'Driver gaze tracking and eyes off the road detection system', IEEE Trans. Intell. Transp. Syst., 16, (4), 2015, pp. 2014–2027

[6] Zheng, R., Nakano, K., Ishiko, H., et al.: 'Eye-gaze tracking analysis of driver behavior while interacting with navigation systems in an urban area', IEEE Trans. Human-Mach. Syst., 46, (4), 2016, pp. 546–556

[7] Jha, S., Busso, C.: 'Analyzing the relationship between head pose and gaze to model driver visual attention', Proc. IEEE Int. Conf. Intell. Transp. Syst. (ITSC), 2016, pp. 1–6

[8] Fridman, L., Lee, J., Reimer, B., et al.: 'Owl and Lizard: Patterns of head pose and eye pose in driver gaze classification', IET Comput. Vis., 10, (4), 2016, pp. 308–313

[9] Tawari, A., Chen, K. H. and Trivedi, M. M.: 'Where is the driver looking: Analysis of head, eye and iris for robust gaze zone estimation', Proc. Int. IEEE Conf. Intell. Transp. Syst., 2014, pp. 988–94

[10] Zhang, X., Sugano, Y., Fritz, M. et al.: 'Appearance-based gaze estimation in the wild', Proc. IEEE Conf. Comput. Vis. Pattern Recogn. (CVPR), 2015, pp. 4511–4520

[11] Wood, E., Baltrusaitis, T., Zhang, X., et al.: 'Rendering of eyes for eye-shape registration and gaze estimation', Proc. IEEE Int. Conf. Comp. Vis. Pattern Recogn. (CVPR), 2015, pp. 3756–3764

[12] Shrivastava, A., Pfister, T., Tuzel, O., et al.: 'Learning from simulated and unsupervised images through adversarial training', Proc. IEEE Conf. Comp. Vis. Pattern Recogn. (CVPR), 3, 2017, pp. 2242–2251.

[13] Ranjan, R., De Mello, S. and Kautz, J.: 'Light-weight head pose invariant gaze tracking', Proc. IEEE Conf. Comp. Vis. Pattern Recogn. (CVPR), 2018, pp. 1–9

[14] Zhang, X., Sugano, Y., Fritz, M., et al.: 'MPIIGaze: Real-world dataset and deep appearance-based gaze estimation', IEEE Trans. Pattern Anal. Mach. Intell., 2017

[15] Liu, W., Anguelov, D., Erhan, D., et al.: 'SSD: Single shot multibox detector', Proc. European Conf. Comput. Vis. (ECCV), 2016, pp. 21–37

[16] Baltrusaitis, T. Morency, L.-P., Robinson, P.: 'Constrained local neural fields for robust facial landmark detection in the wild', IEEE Conf. Comput. Vis. Workshops (ICCVW), 2013, pp. 354–361

[17] Levenberg, K.: 'A method for the solution of certain non-linear problems in least squares', Quarterly of Appl. Math., 2, 1944, pp. 164–168

[18] Marquardt, D.: 'An algorithm for least-squares estimation of nonlinear parameters'. SIAM J. Appl. Math., 11, (2), 1963, pp. 431–441

[19] Broyden, C. G.: 'The convergence of a class of double rank minimization algorithms: 2. The New algorithm', J. Inst. Math. Appl., 6, 1970, pp. 222–231

[20] Fletcher, R.: 'A new approach to variable metric algorithms', Computer J., 13, 1970, pp. 317–322

[21] Goldfarb, D.: 'A family of variable metric methods derived by variational means', Math. Comp., 24, 1970, pp. 23–26

[22] Shanno, D. F.: 'Conditioning of quasi-Newton methods for function minimization', Math. Comp., 24, 1970, pp. 647–650

[23] Haines, E.: 'Point in polygon strategies', in Heckbert, P. (Ed.): 'Graphics Gems IV' (Academic Press, 1994), pp. 24–46

Working towards a Meaningful Transition of Human Control over Automated Driving Systems

Daniël D. Heikoop [1*], Marjan Hagenzieker[1]

[1] Transport & Planning, Civil Engineering, Delft University of Technology, The Netherlands
[*]d.d.heikoop@tudelft.nl

Abstract: Automated vehicles with partial automation, supporting both longitudinal and lateral control of the vehicle, are currently available for the consumer. The consequences of driving with this type of advanced driver assistance systems is not well-known, and could cause the human driver to become out-of-the-loop, or cause other types of adverse behavioural adaptation, leading to dangerous circumstances. Therefore, understanding what the effects of driving with automated driving systems are from the human driver's perspective is becoming imperative. By means of a literature-based approach, this paper presents a framework of human control over automated driving systems. This framework shows the quantified distribution of human behaviour over all the levels of automation. The implications, discrepancies and apparent mismatches this framework elicits are discussed, and recommendations are made to provide a meaningful transition of human control over automated driving systems.

1. Introduction

It is becoming increasingly important to address Human Factors issues with automated driving systems, as consumer vehicles become equipped with exponentially increasing amounts of advanced driver assistance systems that take over parts of the driving task previously performed by the human driver. With the partially automated vehicles (SAE level 2; [1]) already on the road today, both the longitudinal (braking/accelerating, e.g., adaptive cruise control) and lateral (steering, e.g., lane keeping assist) control of the vehicle is being taken over by an automated driving system. Inevitably, this and future technology enabling higher levels of automation will cause out-of-the-loop problems [2], mode confusion [3], and behavioural adaptation [4] issues that need urgent reconsideration in order to maintain safe driving with automated vehicles [5].

Therefore, the transfer of control from the human driver to the automated driving system and vice versa needs to follow a safe and meaningful process that circumvents or even solves the aforementioned issues. The concept of maintaining a form of meaningful human control over automated systems is not new, as it originated from the field of autonomous weapon systems [6]. This concept encompasses all forms of control (i.e., not solely operationally, but also tactically and strategically; cf. [7]) of a human being over an automated system. A recently developed philosophical account defined two conditions that need to be met in order for any system to remain under meaningful human control, namely 'tracking' (i.e., a system should always be able to respond to a human's moral reasons), and 'tracing' (i.e., it should always be possible to trace back how a system came to a decision) [8].

However, in order to be able to attach a meaningful form of control to a human driver—and thus a safe driving behaviour—it is first necessary to assess what behaviour is involved in driving a vehicle (and with automated driving systems), from a human-oriented perspective [9]. Without understanding the full extent of human behaviour within an automated vehicle, it if difficult to know what the notion of 'control' applies to. A taxonomy often used to compare with or extend driver behaviour models from is the taxonomy of Rasmussen [10] (see e.g., [9]).

The taxonomy of Rasmussen [10] distinguishes three levels of human behaviour (explained in more detail in section 2.2) based on the assumption that humans are goal-oriented and thus not mere input-output systems that would structurally adhere to the commands given to them. His assumption encompasses that humans need *reason* (or *meaning*) for a given action, and thus lays the foundation for a human-oriented framework of meaningful human control over automated driving systems.

The question we aim to answer in this paper is: What (types of) human behaviour is involved in automated driving, and to what extent does this behaviour get affected by the introduction of automated driving systems?

In this paper, a quantitative rather than a qualitative approach is taken. Since a quantification of human behaviour with automated driving systems is currently missing, we aim for this approach to serve as a foundation for future research.

2. Development of a framework of human control over automated driving systems

In this literature study, a framework of human control over automated driving systems was developed by means of setting the taxonomy of the SAE related to on-road motor vehicle automated driving systems [1] against the classification of human behaviour determined by Rasmussen [10]. This created a 6x3 framework, entailing 18 fields, each of which to be filled by quantitatively assessing how many driving tasks are subject to each field. The quantitative assessment was done by thorough literature research and, in several occasions for which literature not yet exists, logic and deductive reasoning.

2.1. SAE levels of automation

The levels of automation set out by the SAE are divided into six categories, ranging from level 0 (no automation, or manual driving) to level 5 (full automation). The SAE specifies that these levels are descriptive and technical, rather than normative and legal, meaning that they distinguish these levels by assessing what type of driving task is being taken over by the automated driving systems (e.g., if the execution of steering and acceleration/deceleration is being performed by the automated driving systems, while the monitoring of the driving environment is still to be performed by the human driver, this automated driving systems would be level 2 [partial automation]).

Specifically, the following definitions belong to the six levels of automation:

Level 0: "The full-time performance by the *human driver* of all aspects of the *dynamic driving task*, even when enhanced by warning or intervention systems".

Level 1: "The *driving mode*-specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*".

Level 2: "The *driving mode*-specific execution by one or more driver assistance systems of both steering and acceleration/deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*".

Level 3: "The *driving mode*-specific performance by an *automated driving system* of all aspects of the dynamic driving task with the expectation that the *human driver* will respond appropriately to a *request to intervene*".

Level 4: "The *driving mode*-specific performance by an automated driving system of all aspects of the *dynamic driving task*, even if a *human driver* does not respond appropriately to a *request to intervene*".

Level 5: "The full-time performance by an *automated driving system* of all aspects of the *dynamic driving task* under all roadway and environmental conditions that can be managed by a *human driver*".

2.2. Classification of human behaviour (Rasmussen, 1983)

In his paper, Rasmussen [10] distinguishes three types of human behaviour, namely skill-, rule-, and knowledge-based behaviour. He defines skill-based behaviour as acts or activities which take place without conscious attention or control, and which is automated and highly integrated. Rule-based behaviour is defined as routinely executed acts or activities that follow a stored rule or procedure, often from instruction or preparation. Its distinction from skill-based behaviour depends on the level of training and attention of the person, where skill-based behaviour is unconscious, and rule-based behaviour is consciously based on explicit recollection of facts. Knowledge-based behaviour is the performance of an act or activity during unfamiliar situations, and is goal-controlled. Here, a person needs to plan his/her actions, evaluate those, and consider the best response by functional reasoning. Usually, this is done by selecting from (parts of) previous similar experiences, and piecing together a novel reaction to a novel situation.

2.3. Filling in the blanks: the baseline (SAE level 0)

To set a baseline for the set of skills, rules and knowledge required during (automated) driving, in this paper we consider the case of the driver who recently successfully completed their basic driver training course in a European country. With regard to the choice of this baseline, rather than an 'ideal' or 'average' driver, we believe that these novice drivers are a reasonable baseline for this study, as they represent and express minimal requirements for being allowed to drive a regular vehicle, which would *theoretically* encompass all drivers' skill-, rule-, and knowledge sets.

Therefore, we aimed to find a skillset, laid out by a European organization, which is mandatory to possess in order to acquire a European driving license. This skillset is found to be laid out by the CIECA Road Safety Charter working group's Harmonisation of the Assessment of Driving Test Candidates [11]. This working group identified seven categories of driving skills necessary for passing a driving test, ranging from preparatory skills (e.g., checking the oil level and tyre pressure), via vehicle control (e.g., steering and accelerating/decelerating), to traffic adaptation skills (e.g., merging into traffic), each with their own (sub)categories. A total of 128 unique skills were extracted, which serve as the baseline for driver skill-based behaviour (see Table 1, top left field).

The baseline set of rule-based behaviour was derived from a 1968 convention on road traffic, during which the rules of the road were laid out to increase road safety throughout the European continent, commonly known as the Vienna Convention [12]. In the Vienna Convention [12], 56 articles spread over six chapters discuss everything that enables safe driving in Europe. Excluding some exceptions that are for governmental bodies specifically, the contents of Chapter 2 to 5 are important for every driver to know, which describe the general rules of the road (Ch. 2), and vehicle- (Ch. 3), driver- (Ch. 4), and cycle/moped condition requirements (Ch. 5). Furthermore, since 1968, two important changes have been made in light of the introduction of automated driving systems, namely the inclusion of a new paragraph (5bis) in Article 8, and the amendment of Article 39 [13]. These changes have been included in this paper. The Vienna Convention lists a total of six chapters, in which 37 articles cover 151 main rules that are directly or indirectly related to motor vehicle drivers. In total, these 151 main rules cover 254 unique (sub)rules which form our rule-based behaviour baseline (see Table 1, middle left field). Examples of these rules range from general rules such as that one should not endanger or harm others, and that one should drive on the correct side of the road (left or right, depending on the country one is in), to more complex rules regarding the weight and dimension of goods one can load onto their vehicles, and registration and licensing rules.

2.3.1. The knowledge gap

The third and final step in setting the baseline was finding a set of knowledge-based behaviour for drivers who just received their license. This, however, proved to be no easy task, as this entailed everything else the sets of skills and rules haven't covered yet. Moreover, in search for such a set, the term 'knowledge' needed to be redefined in order to retrieve valuable information, since 'knowledge' as a key search term encompassed too much transient topics. As Rasmussen's [10] definition states this type of behaviour is related to unfamiliar situations, where the driver's behaviour is heavily dependent on the task-capability interaction [14], one can argue this type of behaviour is situationally induced behaviour [15]. Therefore, we aimed to find a set of advanced driver training courses, as those courses aim at training unfamiliar situations. Unfortunately, no such set yet existed. However, some documentation reported several selected countries in Europe [16, 17]. Each reference cited in these documents has been carefully studied, and their results have been summarized (disregarding the results found from non-EU countries; see [17]).

This approach resulted in a set which could be divided into four types of situationally induced behaviours, namely *roadway*-, *traffic*-, *environment*-, and *car*- induced behaviours [15], and totalled 64 unique knowledge-based behaviours one may have to call upon during manual driving as a recently licensed car driver in Europe (see Table 1, bottom left field), such as identifying and recognizing as well as handling under- or oversteer, predictive steering, and defensive driving techniques, such as reciprocation and joint-action.

2.4. Driver Assistance (SAE level 1) and Partial Automation (SAE level 2)

After having set the baseline sets, the effects of the introduction of automated driving systems to human behaviour was assessed. The amount of research done regarding the effects of automation on driver skill is very limited (only works from [18] and [19] were found that were somewhat related), as most research limits itself (understandably) to one or two individual skills like braking or steering. For rule-based behaviour, only the works of [20] were found to be somewhat relevant for this study, so it appeared that a literature-based approach was not warranted hereon forward. Therefore, an inventory of all existing advanced driver assistance systems was sought, and a systems-based approach was taken. This inventory lists six systems that include either longitudinal or lateral assistance, ranging from antilock braking systems to automated parking assistance [18] (see also [21] for a list per vehicle manufacturer). Further investigation found two more variations of such systems, thus totalling eight advanced driver assistance systems currently implemented in consumer market vehicles.

Inspection of these systems regarding their impact on driver skill-, rule-, and knowledge-based behaviour based on the SAE definition, showed that the amount of behaviours required from the driver differs depending on the system that is being used. For example, the autonomous emergency braking system only takes over the skill of making an emergency brake, whereas adaptive cruise control takes over the skill of braking smoothly when a car is in front of you, and several other skills involved in speed adaptation (see [19]). Since the SAE defines level 1 systems to have *either* longitudinal *or* lateral control, the amount of skills required while driving with such a system is flexible. Because driving with advanced driver assistance systems is yet to be included within basic driving courses, no added skills are foreseen as of yet (see Table 1, top second left field).

Regarding the amount of rules a driver needs to adhere to during driving with SAE level 1 systems, we consider the SAE definitions of the levels of automation as added rules to adhere to. Further European legislation regarding automated driving systems are—albeit under development—currently non-existent, although several separate European and non-European countries are progressively adapting rules regarding autonomous vehicles (see e.g., [22]). Next to the additional SAE rules, again, depending on the system in use, varying amounts of rules are being taken over by the advanced driver assistance system. For example, a lane centring system needs to adhere to Article 10, rule 3, concerning the position within a lane, thus making it obsolete for the human driver to adhere to this rule (while driving with that system activated). Adaptive cruise control will, in its turn, need to adhere to Article 13, rule 2, regarding speed limits, and rule 5, regarding the distance between vehicles (see [20]). The results are presented at Table 1, middle second left field.

Lastly, the introduction of novel systems such as advanced driver assistance systems inadvertently introduce novel situations. Thus, in contrast with skill- and rule-based behaviour, these systems will *add* drivers knowledge-based behaviour more than they take over. Although little is known about what situations may occur, several knowledge-based behaviours are expected to be requested by driving with such systems, such as coordinating, cooperating and collaborating with the activated system, but also understanding the distribution of tasks between the driver and the system, as well as knowing when it is safe to engage in secondary tasks [18]. Most of these situations are thus concerned with the new supervisory task of the driver. Note that SAE level 1 systems could potentially take over some knowledge-based behaviour (i.e., a traction control system could take over advanced turn-negotiating techniques, albeit to a limited extent), but this does not outweigh the amount of additional knowledge-based behaviour introduced by these systems. Also note that, especially for novice drivers, the (negative) effect driving with advanced driver assistance systems has on human behaviour is not to be underestimated (see [23]).

With SAE level 2 —or Partial Automation— systems *both* longitudinal *and* lateral control is being taken over by the automated driving system. This could potentially entail a vehicle that has adaptive cruise control with a lane-centring system, or a vehicle that has an automated parking system. Although somewhat dependent of the system, this basically entails that for the human driver the required amounts reach the maximum deviation from the baseline seen at SAE level 1 systems for both skill-, rule-, and knowledge-based behaviour (see Table 1, third left column).

2.5. Conditional Automation (SAE level 3)

From the technical perspective of the SAE, a level 3 automated driving system entails a system that takes over *all* of the *dynamic driving task*. This basically means that all that is left for the human driver to do is to take the necessary preparatory measures before stepping into the vehicle, and drive off automatically. Henceforth, regarding the required amount of skills while driving with a SAE level 3 automated driving system, a massive drop can be foreseen, as none of the skills trained during driver training are called upon, apart from, for example, being able to check the tyre tread and oil level, that the lights still work, and that the mirrors and windows are clean. The entire dynamic driving task, from changing gears to merging in traffic (cf. [19]), will be performed by the automated driving system.

The same applies to the amount of rules the human driver needs to adhere to. Many of the driving-related rules will have to be considered by the automated driving system instead of the human driver, such as the rules regarding overtaking, the priority rules, and rules regarding interacting with vulnerable road users. Nevertheless, a substantial amount of rules are left at the responsibility of the human driver. For example, rules regarding the registration, as well as the loading of your vehicle, and regarding the consequences of disobeying any rule, are still at the human driver's responsibility. Notably, in the event of the vehicle getting involved in an accident—even though the system should be capable of avoiding accidents, as that is essentially part of the dynamic driving task—three rules regarding accident handling will apply to the human driver. Since the automated driving system should be designed to such an extent that an accident should not happen, this situation must be given special attention in the framework (see Table 1, asterisk sign).

When considering the amount of knowledge-based behaviour involved in driving with a SAE level 3 automated driving system, it becomes apparent that this introduces unknown situations to such an extent that quantifying the amount of knowledge-based behaviour required from a human driver is becoming guesswork (see Table 1, question mark sign). Nevertheless, an estimation has been made, based on the SAE's definition of level 3, the consequences of the introduction of automation at SAE levels 1 and 2, and consequences mentioned in [18]. Since most knowledge regarding the dynamic driving task is becoming redundant at this level of automation—as the automated driving system now takes care of that—the amount of knowledge-based behaviour also experiences a decline. What remains are the knowledge-based behaviours regarding car-specific behaviours and understanding one's own behaviour whilst driving (with and without such an automated driving system). However, within this level of automation, one also has to consider the ironies of automation [24], one of which is the deterioration of (unused) skills and rules to a knowledge-based level (see Table 1, exclamation mark sign; see also [18] and [3]).

Up to SAE level 3 automation, the SAE defines that the human driver is expected to serve as a fall-back to perform the dynamic driving task in case of an emergency, like a system malfunction [1] (see Table 1, bold line; see also [25]). This means that for all these levels (SAE levels 0 to 3), the human driver is expected to be able to perform as if (s)he were driving a manual vehicle. Given the considered ironies of automation discussed above, this appears to be misplaced.

2.6. High Automation (SAE level 4) and Full Automation (SAE level 5)

Beyond SAE level 3, where the human driver is still expected to act as a fall-back for safely handling the vehicle in critical situations, most of the quantification of human driver skill-, rule-, and knowledge-based behaviour relies on speculation and debate. Vehicles with SAE level 3 automation don't exist yet, let alone SAE level 4 or 5 [26], however, as with SAE level 3, certain assumptions can be made regarding a human driver's skill-, rule-, and knowledge-based behaviour.

For example, it may be reasonable to assume that with a SAE level 4 automated driving system the human driver will still be responsible for preparing their own vehicle before driving off, while on the other hand not expecting them to still remain in a driving position anymore, making room for other activities such as working on a laptop or reading a book, or even sleeping [27]. Simultaneously, however, one has to wonder how much use a safety belt would still have under such circumstances, or whether people would still actually own their own vehicles, and thus whether or not they still need to be skilled in doing their own safety checks prior to their drive [28].

Where with SAE level 3 automated driving systems the human driver still plays a key (fall-back) role within the driving task, (s)he can be taken completely out-of-the-loop with SAE level 4 automated driving systems. Therefore, certain human driver-oriented rules may not (need to) apply anymore, such as having a physically and mentally fit driver behind the steering wheel, potentially opening the gate for disabled, children and the elderly [29]. As with driver skill-based behaviour, it is however uncertain to what extent certain preparatory rules still apply (e.g., registration rules), while others are likely to still to remain in place (e.g., loading rules). Up to full automation (SAE level 5), it is up to everyone's imagination as to what extent a 'driver' of such a vehicle still needs to abide to a (if any) rule (e.g., will "Don't litter" [Article 7, rule 2] be covered by a fully autonomous vehicle?).

Ultimately, knowledge-based behaviour is unlikely to be part of a driver's task demand while driving a SAE level 4 or 5 automated vehicle, but nevertheless certain situations may occur that places a driver in unknown territory, albeit hard to pin an exact number to that.

Table 1. Framework of human control over automated driving systems. he numbers represent the (range of the) total amount of behaviours that are expected from a novice driver to be present during the respective levels of vehicle automation.

Automation / Human	SAE 0	SAE 1	SAE 2	SAE 3	SAE 4	SAE 5
	No Automation	Driver Assistance	Partial Automation	Conditional Automation	High Automation	Full Automation
Skill	128	127 - 114	114	43	40 - 0?	39 - 0?
Rule	254	255 - 250	250	69* - 66	51 - 29?	29 - 0?
Knowledge	64	64 - 80	80	33?!	0-?!	0?

| Fall-back to human up to SAE 3, means human needs at times adhere to SAE 0 levels.

* In case of accident; i.e., in case the automation is *not* capable of avoiding an accident.

? Higher levels of automation involve unknown situations and definitions.

! Within this stage, driver skill- and rule-based behaviour may already deteriorate to knowledge-based, adding up to a driver's required knowledge-based behaviour.

3. Implications

3.1. The decline in skill- and rule-based behaviour

As can be seen in Table 1, a negative trend in the amount of required skills and rules coincide with the introduction of increasingly autonomous driving systems. With extended exposure to driving with such systems activated, the consensus is that an actual loss of skill can be expected (e.g., [19, 24, 30]). Only by consistent maintenance of these skills, and rehearsal of these rules, one could avoid having these deteriorate to a knowledge-based behaviour level (cf. exclamation mark sign at Table 1), but that requirement simultaneously beats the purpose of automated driving systems altogether, as these systems—as goes for many other automated systems—are predominantly there to replace the human as the operator [31].

3.2. The rise and fall of knowledge-based behaviour

Contrary to the trend seen with skill- and rule-based behaviour, knowledge-based behaviour first experiences a rise in requests for the human driver. The introduction of advanced driver assistance systems appears to introduce more novel situations than that they dissolve. Behavioural changes such as, but certainly not limited to, becoming complacent and having to supervise an automated system will have to be accounted for in order to ensure safe driving with such systems (e.g., [32, 33, 34]).

Only during SAE level 3 automated driving we begin to see a decline in the request for knowledge-based behaviour, which is because of the execution of "all aspects of the dynamic driving task" by an automated driving system [1]. However, since the SAE also states that they have "the expectation that the *human driver* will respond appropriately to a *request to intervene*", at least the behavioural changes mentioned above are to be expected to become of importance to a driver's knowledge-based behaviour. To what extent a *request to intervene* requires knowledge-based behaviour is yet to be determined, but quick regeneration of awareness of the situation at hand is considered to be one of the requirements (e.g., [35, 36, 37]).

3.3. The human driver as a fall-back mechanism

As mentioned in section 2.5, the human serving as a fall-back in case of emergency appears misplaced. At the stage where a person has been driving with a SAE level 3 automated driving system for extended periods of time, reclaiming the wheel may be futile as the majority of skills, rules and knowledge necessary for safe driving have not been mobilized in this time (see Table 1). Especially when this level of automated driving encompasses novel techniques such as platooning, more exacerbating behavioural adaptations may occur, such as carryover effects [38], and loss of task engagement [39], to name a few. Given the fact that a deviation in skill-, rule- and knowledge-based behaviour from manual driving occurs throughout all levels of automation, it appears paramount to reconsider the driver's role as a fall-back mechanism during automated driving, especially when given the time to 'forget' about their learned skills and rules (see also e.g., [18] and [24]) .

3.4. SAE level 4 and 5 automation: the path of the unknown

Automated vehicles of SAE levels 4 and 5 are currently only things of the future. Therefore, little knowledge exists on what the effects of those automated driving systems would be. One thing is clear though, and that is that the human will become completely removed from the driving task. Based on the framework presented at Table 1, we have to assume that at this stage, the driver is (almost) completely incapable of resuming manual control, so even a gradual decrease in the level of automation could potentially have disastrous consequences. From this, it appears that the fall-back threshold up to SAE level 3 (bold line at Table 1) has become a point-of-no-return, in the sense that manual intervention is not expected according to its SAE definition, but also not possible anymore.

This does not mean, however, that by having taken into account all of the dynamic driving tasks by the automated driving system it has achieved an infallible machine. It also implies that the as of yet unforeseen newly introduced situations that come with these new type of automated driving systems have to be taken into account (cf. Table 1, two bottom right fields). To give the reader some examples of what might be laying in the autonomous driving future, see [40] and [41].

Lastly, regarding new legislation to be set out by European legislator bodies, the new situations that will arise also requires a legal safety system design. Example suggestions for applicable rules for the new type of driving with automated systems are presented in [20].

4. Limitations, Recommendations, and Future Research

4.1. Limitations of this research

This research attempted to develop a framework of human control over automated driving systems by *quantitatively* assessing the effects various levels of automation has on human behaviour. This means that the framework presented in Table 1 does not provide answers about the effects on the *quality* of human behaviour. It can be argued that certain skill-, rule-, or knowledge-based behaviours have more weight than others in the driving task.

Another limitation of this research is that although a literature-based quantitative approach was attempted in this study, not all fields in the framework were viable for this approach, given the futuristic nature of the higher levels of automation (e.g., SAE level 4 and 5). The actual numbers may be completely different when actual SAE level 4 and 5 automated driving systems exist.

A third limitation is that the framework is not empirically tested. Although validated by thorough literature research, empirical testing of the framework could provide more insights into its validity.

The final point of discussion that should be made here is that the adopted classification of human behaviour of Rasmussen [10] is not the only suitable, nor necessarily the best classification that could be used for the development of such a framework. Examples of similar classifications of human behaviour are the Markov dynamic model of driver action [42], the conceptualisation of a driver's task [43], or the hierarchical structure of the road user task [7]. Michon [7] further summarizes several more in-depth models of human behaviour (see also [44] and [45]). Although the classification used in this paper provided valuable insights that could help increase safety in driving with automated driving systems, we will not discourage attempts of frameworks with different categorisations, as those could potentially point out other bottlenecks and design issues related to human behaviour.

4.2. Mismatch between supply and demand

The developed model sheds light on a serious problem with respect to the role a human driver is supposed to play within an automated driving system. At various levels of automation, large deviations from manual driving concerning skill-, rule-, and knowledge-based behaviour raises issues regarding what we still can and still are supposed to do (cf. [23]). The apparent mismatch between the availability of skills, rules and knowledge at especially the higher levels of automation, and what is requested from the driver (e.g., acting as a fall-back) suggests that the current transfer of control within an automated driving system needs an overhaul, and, more importantly, a (meaningful) human-oriented transfer of control.

Important to note is that the issue with the transfer of control is not only the mismatch between supply and demand, but also the possibility of mode error if this transfer is not communicated appropriately [3].

4.3. Future research

The developed framework presented in this paper suggests a human-oriented taxonomy of levels of automation, in order to secure a safe and meaningful transfer of control. Future research should investigate how such a human-oriented taxonomy could look like.

Next to empirically testing the validity of the framework presented here, it is suggested to have the to be developed human-oriented taxonomy empirically tested too.

Furthermore, predictive models like those used in economics or econometrics, or those used in the estimation of logistics- and fuel consumption benefits of platoons (see e.g., [46]), could be used to attempt more sound calculations of the effects of the higher, futuristic, levels of automation.

Lastly, a qualitative approach could be made regarding a framework that assesses the effects of automated driving on human behaviour.

5. Acknowledgements

This study is funded through the NWO-sponsored project, called Meaningful Human Control over Automated Driving Systems [Project Number = 313-99-329].

6. References

[1] SAE International. (2016). *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles*. Washington, DC: SAE International.

[2] Kaber, D. B. & Endsley, M. R. (1997). Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Safety Process*, *16*, 126-131.

[3] Stanton, N. A., & Marsden, P. (1996). From fly-by-wire to drive-by-wire: Safety implications of automation in vehicles. *Safety Science*, *24*, 35-49. doi: 10.1016/S0925-7535(96)00067-7

[4] Rudin-Brown, C. M. & Parker, H. A. (2004). Behavioural adaptation to adaptive cruise control (ACC): implications for preventive strategies. *Transportation Research Part F: Traffic Psychology and Behaviour*, *7*, 59-76. doi: 10.1016/j.trf.2004.02.001.

[5] Saffarian, M., De Winter, J. C. F., & Happee, R. (2012). Automated driving: Human-factors issues and design solutions. *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting*, 2296-2300. doi: 10.1177/1071181312561483.

[6] Future of Life Institute. (2015). Autonomous Weapons: An open letter from AI & Robotics Researchers. from https://futureoflife.org/open-letter-autonomous-weapons/.

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

[8] Santoni de Sio, F. & Van de Hoven, J. (2018). Meaningful Human Control over autonomous systems: A philosophical account. *Frontiers in Robotics and AI, 5*, 1-14. doi: 10.3389/frobt.2018.00015.

[9] Ranney, T. A. (1994). Models of driving behavior: A review of their evolution. *Accident Analysis and Prevention, 26*, 733-750. doi: 10.1016/0002-4575(94)900051-5.

[10] Rasmussen, J. (1983). Skills, rules, and knowledge; Signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, *13*, 257-266.

[11] CIECA Road Safety Charter working group (2006). *Harmonisation of the Assessment of Driving Test Candidates.* Berlin, Germany.

[12] United Nations. (1968). *Convention on road traffic*. Vienna, Austria.

[13] United Nations. (2014). *Report of the sixty-eighth session of the Working Party on Road Traffic Safety*. Geneva, Austria. 24-26 March.

[14] Fuller, R. (2004). Towards a general theory of driver behaviour. *Accident Analysis and prevention, 37*, 461-472. doi: 10.1016/j.aap.2004.11.003.

[15] McKnight, A. J., & Adams, B. B. (1970). Driver education task analysis. Volume II: Task analysis methods. Alexandria, VA: Human Resources Research Organization.

[16] SWOV (2009b). *Voortgezette rijopleiding voor beginners*. SWOV-Factsheet, april 2009. Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV, Leidschendam.

[17] Lynam, D., & Twisk, D. (1995). *Car driver training and licensing systems in Europe.* Transport Research Laboratory. Crowthorne, Berkshire.

[18] Spulber, A. (2016). *Impact of automated vehicle technologies on driver skills*. Ann Arbor, MI: Center for Automotive Research.

[19] Young, M. S., & Stanton, N. A. (2007). What's skill got to do with it? Vehicle automation and driver mental workload. *Ergonomics, 50*, 1324-1339. doi: 10.1080/00140130701318855

[20] Vanholme, B., Gruyer, D., Lusetti, B., Glaser, S., & Mammar, S. (2013). Highly automated driving on highways based on legal safety. *IEEE Transactions on Intelligent Transportation Systems, 14*, 333-347. doi: 10.1109/TITS.2012.2225104

[21] "Which cars have self-driving features for 2017?" (2017). from https://www.cars.com/articles/which-cars-have-self-driving-features-for-2017-1420694547867/

[22] Wolfers, B. (2017). Selbstfahrende Autos: Ist das erlaubt? Einführung in die Regulierung des automatisierten Fahrens und den StVG-Änderungsentwurf der Bundesregierung von Januar 2017. In M. Schlamp (Ed.) *RAW, Recht, Automobil, Wirtschaft*, (pp. 2-13). DFV Mediengruppen, Frankfurt am Main.

[23] Tsapi, A. (2015). *Introducing Advanced Driver Assistance Systems (ADAS) into drivers' training and testing: The young learner drivers' perspective.* (Master's thesis), Delft University of Technology.

[24] Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*, 775-779.

[25] "Updated: Autonomous driving levels 0 to 5: Understanding the differences". (2016). from https://www.techrepublic.com/article/autonomous-driving-levels-0-to-5-understanding-the-differences/

[26] "What are autonomous car levels? Levels 1 to 5 of driverless vehicle tech explained. (2018). from https://www.carmagazine.co.uk/car-news/tech/autonomous-car-levels-different-driverless-technology-levels-explained/

[27] "Drivers of BMW iNext will be able to sleep behind the wheel". (2017). from https://www.motorauthority.com/news/1110737_drivers-of-bmw-inext-will-be-able-to-sleep-behind-the-wheel

[28] Stocker, A., & Shaheen, S. (2017). *Shared automated vehicles: Review of business models.* Berkeley, CA: International Transport Forum.

[29] Milakis, D., Van Arem, B., & Van Wee, B. (2017). Policy and society related implications of automated driving: A review of literature and directions for future research. *Journal of Intelligent Transportation Systems, 21*, 324-348. doi: 10.1080/15472450.2017.1291351.

[30] Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans, 30*, 286-297. doi: 10.1109/3468.844354.

[31] Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 39*, 230-253. doi: 10.1518/001872097778543886.

[32] Brookhuis, K. A., De Waard, D., & Janssen, W. H. (2001). Behavioural impacts of Advanced Driver Assistance Systems—An overview. *European Journal of Transport and Infrastructure Research, 1*, 245-253.

[33] Lindgren, A., & Chen, F. (2006). State of the art analysis: An overview of Advanced Driver Assistance

Systems (ADAS) and possible Human Factors issues. *Human Factors and Economics Aspects on Safety*, 38-50.

[34] Stanton, N. A., & Young, M. S. (2005). Driver behaviour with adaptive cruise control. *Ergonomics, 48*, 1294-1313. doi: 10.1080/00140130500252990.

[35] De Waard, D., Van der Hulst, M., Hoedemaeker, M., & Brookhuis, K. A. (1999). Driver behavior in an emergency situation in the automated highway system. *Transportation Human Factors, 1*, 67-82. doi: 10.1207/sthf0101_7.

[36] Eriksson, A. & Stanton, N. A. (2017). Takeover time in highly automated vehicles: Noncritical transitions to and from manual control. *Human Factors, 59*, 689-705. doi: 10.1177/0018720816685832.

[37] Lu, Z., Coster, X., & De Winter, J. C. F. (2017). How much time do drivers need to obtain situation awareness? A laboratory-based study of automated driving. *Applied Ergonomics, 60*, 293-304. doi: 10.1016/j.apergo.2016.12.003.

[38] Skottke, E.-M., Debus, G., Wang, L., & Huestegge, L. (2014). Carryover effects of highly automated convoy driving on subsequent manual driving performance. *Human Factors, The Journal of the Human Factors and Ergonomics Society, 56*, 1272-1283. doi: 10.1177/001872081452594.

[39] Heikoop, D. D., De Winter, J. C. F., Van Arem, B., & Stanton, N. A. (2017). Effects of platooning on signal-detection performance, workload, and stress: A driving simulator study. *Applied Ergonomics, 60*, 116-127. doi: 10.1016/j.apergo.2016.10.016.

[40] Chipchase, J. (2016). Driver behaviours in a world of autonomous mobility. from https://medium.com/studio-d/15-more-concepts-in-autonomous-mobility-8fd1c794e466

[41] Chipchase, J. (2014). Twelve concepts in autonomous mobility. from https://medium.com/studio-d/concepts-in-autonomous-mobility-80732bc4a44d

[42] Pentland, A., & Liu, A. (1999). Modeling and prediction of human behavior. *Neural Computation, 11*, 229-242.

[43] De Winter, J. C. F., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation Research Part F: Traffic Psychology and Behaviour, 27*, 196-217. doi: 10.1016/j.trf.2014.06.016.

[44] Stanton, N. A., & Young, M. S. (2000). A proposed psychological model of driving automation. *Theoretical Issues in Ergonomics Science, 1*, 315-331. doi: 10.1080/14639220052399131.

[45] Heikoop, D. D., De Winter, J. C. F., Van Arem, B., & Stanton, N. A. (2016). Psychological constructs in driving automation: A consensus model and critical comment on construct proliferation. *Theoretical Issues in Ergonomics Science, 17*, 284-303. doi: 10.1080/1463922.X.2015.1101507.

[46] Janssen, R., Zwijnenberg, H., Blankers, I., & De Kruijff, J. (2015). Truck platooning: Driving the future of transportation.

Predictors and Risk Perceptions of Using Cell Phones while Driving among Young Adult Drivers

Scarlett Mia S. Tabuñar [1]*

[1] Department of Emergency Medicine, University of the Philippines-Philippine General Hospital, UP-College of Medicine, Philippines
*sstabunar@up.edu.ph, miatabunar@hotmail.com

Abstract: The Philippines is expecting a rise in the number of drivers that use mobile phones while driving. It is known as the 'texting capital of the world'. The objectives of this study were to determine the predictors, risk perceptions and the prevalence of cell phone use while driving among trainee residents of the University of the Philippines-Philippine General Hospital. This cross-sectional study employed total enumeration. A survey was first distributed to the target population, followed by a focus group discussion. Chi-square and multiple logistic regression were used to analyse data. Included in the final analysis were 175 drivers aged 25-30 years (mean=27.90 $\pm$1.34). There was no significant difference in the risk perceptions of cell phone users vs. non-users, and most perceived hands-free devices safer to use (p=0.030). The reported prevalence is 90.68%; drivers have a significant overall unsafe attitude (p=0.007), and an unsafe attitude when using handsets when driving, even when this is known to be dangerous (p=0.003). In conclusion, driving with hands-free devices was perceived to be safer, although drivers have a high overall unsafe attitude. Driving for more than 2 years and having an unsafe attitude were found to be significant predictors of cell phone use while driving. Countermeasures must take into account these factors when instituting behavioural modification strategies and road safety policies concerning unsafe and distracted driving.

KEYWORDS: attitude, cell phone use while driving, driving experience, distracted driving, risk perceptions, young adult drivers

1. Introduction

Distracted driving (DD) is one of the key factors cited by the World Health Organization (WHO) that needs to be addressed by governments in order to prevent road traffic injuries (RTI) [1]. The Centers for Disease Control and Prevention (CDC) defines distracted driving as any activity that takes a person's attention away from the primary task of driving. Activities such as using cell phones, texting, utilizing navigation technologies (GPS) and eating are all considered distractions which could endanger road users' safety [2]. There are three main forms of distraction while driving: manual, visual and cognitive. Manual distraction involves taking one's hands off the steering wheel while visual distraction occurs when the driver's eyes are taken off the road. Cognitive distraction, on the other hand, happens when the individual's focus is not directly on the act of driving, causing his or her attention to wander [2]. Of particular public health interest, due to the advent of technology, is texting and talking on mobile phones while driving, as they pose the most significant and real danger by combining all three types of DD [3]. Several studies show that the distraction caused by hand-held phones could impair driving performance, e.g. longer reaction times (notably braking, and reaction time, also reaction to traffic signals), impaired ability to keep in the correct lane, and shorter following distances, resulting in overall reduction in awareness of the driving situation [4]. Cell phone use while driving increases the likelihood of a road crash by four-fold [5]. Simulation studies report that this type of distraction could cause a similar decrement in driving performance to a person with a 0.8 percent blood alcohol level, the point at which drivers are generally considered intoxicated [6]. Hands-free devices (e.g. earphones, speaker-phones, Bluetooth, etc.) do not appear to minimize the deleterious effects of DD, as evidence reveals that hands-free cell phone road users execute the tasks of driving with the same diminution in ability as those who do not [7], [8].

The 2014 National Highway Traffic Safety Administration (NHTSA) report on distracted driving estimates that 71% of teens and young drivers compose and send text messages, and 78% read short message services [9]. An alarming statistic reveals an increase of drivers text-messaging or visibly manipulating handheld devices from 1.7% in 2013 to 2.2 % in 2014, with young drivers (age 16 to 24) using electronic devices at higher rates [9]. Distracted driving, particularly through mobile phone use, is much more common among young adult drivers (under 30 years of age) [10]. In the Philippines the National Statistics Coordination Board (NSCB) reports that from 2001-2006 the highest spike in the cause of road traffic crash of more than five times is due to cell phone use, which ranked 12th amongst the most common cause of traffic accidents in 2006 [11]. More than 70 countries worldwide enforce restrictions and bans on the use of mobile phones while driving [12], it was only on July 2015 that the Philippines enacted the Republic Act 10913, or Anti Distracted Driving Act, defining and penalizing distracted driving. Under the new law, "distracted driving" is defined as "using a mobile communications device to write, send, or read a text-based communication or to make or receive calls," and "using an electronic entertainment or computing device to play games, watch movies, surf the Internet, compose messages, read e-books, perform calculations, and other similar acts" [13]. As most of the evidence for distracted driving comes from research performed in industrialized countries, there is a dearth of local literature investigating the extent of the problem, particularly in young adult drivers. It is therefore the aim of this study to determine the predictors, risk perceptions and the prevalence of distracted driving among young doctors training at the University of the Philippines-Philippine General Hospital (UP-PGH), aged under 30 years, who use cell phones, for both text-messaging and conversing while driving.

2. Methodology

The study was undertaken in two stages: first, survey questionnaires were distributed to the target population of young adult drivers followed by a focus group discussion (FGD), with one of the identified resident groups via convenient sampling who were available for the activity. The main purpose of which was to gather a more detailed information on the topic.

The study design was cross-sectional with tool questionnaires given to all year levels (1st to 5th year) of the residents at UP-PGH. Total enumeration was employed in order to capture the subset of drivers in the study population. The trainee residents were chosen for their ages, which fall within the 24-30 year-old range, well within the age group of interest. It was conducted over a two month period, from July to August 2017. The total number of residents hired by UP-PGH at the start of the year was 533 according to the Human Resource department of the hospital. Non-drivers, drivers aged 31 and above, and those who did not agree to participate in the study, were excluded. The calculated sample size (n) had a confidence level of 95% (Z 1- $\alpha/2$ =1.96), with expected proportion (p) = 0.5, and absolute precision (d) of 0.04 and alpha=0.05 is 306. Ethics approval came from UP-Manila Research Ethics Board (UPM-REB-2017-149-01), which was secured prior to the start of the study, and informed consent for the survey was waived, as the self-administered questionnaire was anonymized in order to protect the privacy of respondents. However, it was accordingly secured for the FGD.

The structured questionnaire was developed based on the objectives of the study, a review of the related literature [12], [14], and constructed in a way that is more apt to the local setting. It consisted of four sections, namely socio-demographic profile, risk perceptions, distracted driving behaviour survey and attitude toward distracted driving. It was distributed by a research assistant to the residents either during one of the departmental conferences, or at any preferred time of their convenience. The socio-demographic profile part had five questions and an added item of inquiry about the knowledge of the Anti-Distracted Driving Law penalizing the act. The section on risk perceptions had four questions on the use of hands-free devices, the dangers of cell phone use that can result in collisions, and cell phone use being just as dangerous as alcohol-impaired driving. Responses were based on a five-point Likert scale with 1=Strongly Disagree, 2= Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree. In order to differentiate the perceptions of risk, responses were collapsed into to two categories: safe risk perception was defined as Likert Scales that agreed with statements complying with established national laws on distracted driving. According to Distracted Driving Act (RA 10913), a motorist engaging in any of the following acts in a motor vehicle in motion or temporarily halted at a red light, whether diplomatic, public or private, is considered unlawful; (a) using a mobile communications device to write, send or read a text-based communication or to make or receive calls, and other similar acts; and (b) using an electronic entertainment or computing device to play games, watch movies, surf the internet, compose messages, or read e-books[13]. Those responses, under the Likert Scale that were contrary to the Distracted Driving Act, including 'neutral' answers, were considered unsafe risk perceptions.

The distracted driving survey focusing only on cell phone use while driving was adopted and modified from a version of the 11-item Distracted Driving Survey of Bergmark et al. It is a validated tool measuring cell phone-related distracted driving for drivers age 24 and below [15]. This section has four questions (4) concerning cell phone and hands-free device use, and asked whether the resident had used their mobile phone to view other mobile phone applications such as maps, directions and social media while driving during the previous 30 days; the response was binary, recorded either yes or no.

Finally, five (5) questions on attitudes towards cell phone use during driving were modelled after the questions used by Harrison in order to evaluate college students' perceptions of text messaging while driving [14]. The response was similar to that for risk perception using the same description of a 5-point Likert Scale, and interpretation was similarly divided into two groups: safe attitudes (Likert Scales in agreement with DD Laws) and unsafe attitudes (Likert Scales, including 'neutral,' that went against DD Laws).

The FGD topics were guided by several reports from countries that have extensively studied and made progress in addressing risky driving behavior, particularly distracted driving by using handphones [16], [17], [11]. The principal investigator conducted the FGD among residents of the Department of Emergency Medicine (DEM). The FGD explored the issues and constructs included in the structured survey.

Socio-demographic data and qualitative data were encoded in Microsoft Excel and analysed using STATA V12 . A summary of the descriptive data was tabulated through graphic presentation. Chi-square was used to determine the associations between the variables of interest. A univariate comparison was performed on the socio-demographic data, risk perceptions and attitudes, which identified the significant variables ($p<0.05$). Multiple logistic regression was then utilized to ascertain the predictors of cell phone use while driving. An odds ratio with a 95% confidence interval was used as the summary statistics.

3. Results

A total of 393 residents answered the survey out of the 533 residents but only 175 drivers (44.52%) aged 25-30 years were included in the analysis, which satisfied the inclusion criteria. The mean age of the driving respondents was 27.90 $\pm$ 1.34, the youngest being 25 and the oldest 30. More than half (54.29%) were men and 52.98% were in the combined mid-range annual family income of P100, 001 to P1 million (~USD 1,935 to USD 19,357). One hundred and two doctors (58.96%) admitted being involved in a road traffic accident (RTA), mostly as drivers (42.86%), while 26.37% were involved as passengers, and 30.77% as both. The socio-demographic profile of the driving residents is summarized in Table 1.

Table 1. Socio-demographic Profile of Drivers

Data		***n= 175***	
1. Age (years)	Mean	27.90 $\pm$ 1.34 (min= 25; max= 30)	
		Number	%
2. Sex	Male	95	54.29
	Female	80	45.71
3. Annual Family Income (USD 1=Php 51.66)*			
	P100,000 and less (<USD1,935)	16	9.52
	P100,001 to P500,000 (USD1,935- 9,679)	44	26.19
	P500,001 to P1,000,000 (USD-9,679-19,357)	45	26.79
	P1,000,001 and above (>USD 19,357)	63	37.50
4. Involvement in RTC			
	No	71	41.04
	Yes	102	58.96
	As driver	39	42.86
	As passenger	24	26.37
	Both	28	30.77
5. Driving for how many years?			
	≤ 2	25	14.45
	> 2	148	85.55
6. Do you know that distracted driving is penalized under the "anti-distracted driving" law?			
	Yes	165	94.29
	No	10	5.71

*Conversion rate of US dollar to Philippine peso as 28 Oct 2017

.

Table 2.Frequency of and Risk perception of Cell Phone Use while Driving; **Cell Phone Users vs Non-Users**

Items	DRIVERS Using Cellphone				Drivers NOT Using Cellphone				
	Safe Risk Perception (Likert Scales in agreement with Anti-DD Laws)		Unsafe Risk Perception (Likert Scales contrary to Anti-DD Laws including neutral)		Safe Risk Perception (Likert Scales in agreement with Anti-DD Laws)		Unsafe Risk Perception (Likert Scales contrary to Anti-DD Laws including neutral)		*p-value*
	n	%	n	%	n	%	n	%	
1. Hands-free devices are safe to use when driving	16	11.03	129	88.97	5	33.33	10	66.67	***0.030***
2. Cellphone use is NOT always dangerous while driving	85	58.22	61	41.78	12	80.00	3	20.00	0.101
3. Cellphone use will more likely result in a road crash / collision?	118	81.38	27	18.62	14	93.33	1	6.67	0.473
4. Cellphone use is as dangerous as alcohol-impaired driving?	88	60.27	58	39.73	9	60.00	6	40.00	0.984
OVERALL Risk perception of Distracted Driving	13	9.03	131	90.97	4	26.67	11	73.33	0.059

Regarding driving experience, 85.55% had been driving for more than two years, and a considerable percentage (94.29%) of respondents knew that distracted driving is penalized under the anti-distracted driving law.

The overall risk perception of mobile phone use during driving had no significant findings amongst either users or non-users. However, more drivers who used cell phones perceived using hands-free devices to be safer (p=0.030) (Table 2). Results showed that a considerable proportion of residents (65.22%) either sent or read a texts, called or answered a call while, while a more significant percentage (84.47%) accessed their handsets to view maps, directions or navigation applications. More than half (55.90%) of respondents used hands-free devices such as earphones, speakerphones, Bluetooth devices, etc., while behind the wheel and 49.69% viewed and read messages on social media sites via their phone while driving (Table3).

Not all respondents answered all the questions resulting in missing values on a number of items; the proportion of missing data is relatively small thus they were omitted in the final analysis. Overall reported cell phone usage was 146 or 90.68% out of the only 161 residents who registered a response on self-reported cell phone use. The mean age was 27.39 $\pm$ 1.34, with more males (56.85%) who engaged on this distracting activity, while 40% had an annual family income of more than P1 million (~ >USD 19,357). More than half or 59.72% were involved in a road traffic crash mostly as a driver (42.67%), and a considerable percentage (87.59 %) were driving for more than two years. Only 7 (4.79%) of the 146 cell phone users admitted not knowing the implementation of the ‘Anti-Distracted Driving Law’ (Table 4). Only driving experience of more than two years (p=0.002) had a significant association with handheld phone use while driving among the study participants.

Table 3. Frequency of drivers who use cell phones while driving

Items In the Last 30 days?	Yes		No	
	n	%	n	%
1. Did you use your cellphone while driving (including texting, reading text, calling or receiving call)?	105	65.22	56	34.78
2. Do you use hands-free devices when using your cellphone while driving? (eg. earphones, speaker phone, Bluetooth)	90	55.90	71	44.10
3. Have you used your cellphone to view maps, directions or navigation apps while driving? (e.g. google map, Waze, GPS etc.)	136	84.47	25	15.53
4. Have you used your cellphone to view or read messages on social apps or sites while driving? (e.g. Facebook, Twitter, Instagram, Snapchat etc)	80	49.69	81	50.31

Table 4. Association of drivers who use cell phones while driving, to a socio-demographic profile

Data		*n= 146*		
1. Age	*Mean:* 27.39 ± 1.34 (min= 25; max= 30)			
		n	%	p-value
2. Gender	Male	83	56.85	0.081
	Female	63	43.15	
3. Annual Family Income (USD 1=Php 51.66)	P100,000 and less (<USD1,935)	11	7.86	0.241
	P100,001 to P500,000 USD1,935- 9,679)	35	25.00	
	P500,001 to P1,000,000 (USD-9,679-19,357)	38	27.14	
	P1,000,001 and above (>USD 19,357)	56	40.00	
4. Involvement in RTC	Yes	86	59.72	0.632
	As driver	32	42.67	
	As passenger	19	25.33	
	Both	24	32.00	
	No	58	40.28	
7. Driving for how many years?	≤2	18	12.41	***0.002***
	>2	127	87.59	
8. Do you know that distracted driving is penalized under the 'Anti-Distracted Driving Law'?	Yes	139	95.21	0.199
	No	7	4.79	

Comparative analysis showed that drivers who engaged in this type of distracted activity had significantly higher overall unsafe attitudes vis-a-vis to those who did not (p=0.007), and the same result was noted for unsafe attitudes of those using handsets, even when the drivers knew it was dangerous to do so while driving a vehicle (p=0.003) (Table 5).

The preliminary results of the univariate logistic regression analysis revealed risk perception (p=0.046), years of driving (p=0.001) and attitude (p=0.005) as possible predictors of cell phone use while driving (See Appendix, Table 6 for the univariate logistic regression results).

Model building using multiple logistic regression that sequentially omits the variable that had the highest p-value identified predictors. The final model showed that attitude and more than two years of driving as significant predictors of cell phone use while driving. (See Appendix, Table 7 for multiple logistic regression results).

3.1 Key Findings of the Focus Group Discussion

The principal investigator facilitated the FGD via convenient sampling among residents of the Department of Emergency Medicine (DEM) following one of their academic activities. Topics included in the discussion were risk perception, attitudes and the socio-demographic profile that influenced handheld phone use in young adult drivers. A more in-depth approach was employed in order to gain more specific and detailed information. The objective of the discussion was to acquire untapped insights into distracted driving, particularly concerning cell phone use. It also aimed to gather sentiments, inputs, and opinions that may not have been captured by the questionnaire. The discussion clarified some of the responses in the preliminary survey and identified salient factors relevant to the subject, which helped to analyse the results in a meaningful way.

Seven DEM residents consented to participate after an initial explanation of the intent of the activity; the group consisted of four males and three females belonging to all year levels of training. The whole discussion was audio-taped and lasted for 57 minutes. It was a free-flowing discourse of ideas expressed in both the vernacular and in English.

3.1.1 Driving Background of FGD participants: The driving history of the participants ranged from two to twenty years, driving motorcycle and cars, from 7 to 10 times per week and with most of them driving in a city traffic environment. Most admitted to have once been involved in a road traffic crash (RTC); five as drivers, one as a passenger and one as both; no one received a police citation or traffic infraction.

Table 5. Frequency of and Attitude towards Cell Phone Use while Driving**; Cell phone Users vs Not Users**

	DRIVERS Using Cell phone				Drivers NOT Using Cell phone				
	Safe Attitude (Likert Scales in agreement with anti-DD Laws)		Unsafe Attitude (Likert Scales contrary to anti-DD Laws including neutral)		Safe Attitude (Likert Scales in agreement with anti-DD Laws)		Unsafe Attitude (Likert Scales contrary to anti-DD Laws including neutral)		*p-value*
	n	%	n	%	n	%	n	%	
1. It is unsafe to use a cell phone while driving? (including texting, reading texts, calling or receiving calls	114	78.08	32	21.92	13	86.67	2	13.33	0.740
2. It should be illegal to use a cell phone while driving	60	41.10	86	58.90	10	66.67	5	33.33	0.057
3. Using a cell phone while driving can be dangerous, so I won't do it	68	46.58	78	53.42	13	86.67	2	13.33	***0.003***
4. It is only me who will be affected if I want to text while driving	116	79.45	30	20.55	14	93.33	1	6.67	0.306
5. Using a cell phone while driving is not always distracting	52	35.62	94	64.38	9	60.00	6	40.00	0.064
OVERALL Attitude on using cell phone while driving	22	15.07	124	84.93	7	46.67	8	53.33	***0.007***

3.1.2 Socio-demographic differences of risk perceptions on cell phone use while driving: The insights gathered were that younger, male drivers with higher educational attainments and higher annual family incomes engaged more in distracted driving than their counterparts. Older and more experienced drivers were more careful, focused and vigilant when driving, knowing well that their reflexes were not as fast as when they were younger. Likewise, their years of experience taught them to be less distracted and non-complacent when driving. On the other hand, younger, less experienced drivers were more 'techie,' over-confident and agitated when driving. Young drivers used their mobile phones during long drives to prevent them from falling asleep on the road.

Male drivers were more prone to doing more things while driving, eg. tinkering on the car stereo, compared to female drivers. Men were also more distracted and inattentive when driving, leading to difficulty in manoeuvring, when mistakes in spatial estimation occurred. Moreover, drivers with higher educational backgrounds and higher annual family incomes have a higher likelihood of possessing telecommunication devices, and are also likely to have more things preoccupying their minds, whether personal or occupation-related, resulting in greater cell phone use when driving. Drivers in the higher income bracket were obliged to answer work-related messages or calls.

3.1.3 The prevalence of distracted driving particularly through cell phone use while driving: This was high among the group, with or without hands-free devices, due to the utilization of navigational apps, e.g. Waze, maps etc.

3.1.4 Reasons for cell phone use when driving: Calls were usually answered or made to significant persons e.g. girlfriends, parents or workmates, with no particular preference whether the vehicle was moving or stationary. One interesting insight was that this was thought to be more prevalent in the younger generation of drivers, because they have acquired a 'reflex' for automatically answering ringing cell phones in any situation, whether driving or not. Millennials in particular have answered their phones the instant they rang, or read a message as soon as it arrived, a habit which was unconsciously imbibed even while driving.

One example of when cell phones will be avoided during driving was when in-laws are in the vehicle to make them feel safe. Driving up a steep incline, or dangerous road conditions, e.g. down the side of a mountain, were just a few of the situations when cell phones were not used.

3.1.5 Risk perceptions of cell phone use while driving: The most dangerous consequence of distracted driving is of being involved in a road traffic crash, a fact recognized by most in the group. Environmental or weather conditions, experience, and driving skill are all relevant to the potential hazards when engaging in distracted driving. Inclement weather or poorly lit roads could impair the vision of the driver, and the inherent risk and probability for an accident could even be higher for newly qualified drivers.

3.1.6 Attitudes towards cell phone use while driving: Being aware of the probable danger of distracted driving did not deter most of the drivers from using their handsets. The majority did not use their phones when driving with a significant person in their lives, such as spouse, parents or children. Designating another passenger to answer phone calls or texts was one safety strategy. However, it was sometimes inevitable to use the navigation apps.

3.1.7 Opinions on the law on "Anti-Distracted Driving" and its enforcement: All agreed that drivers engaging in mobile phone use when driving should be penalized, but all also noted that the current laws on this risky behavior were quite lax in their implementation, and enforcement was not as strict as it should be. A higher and a stiffer penalty should be imposed if violations result in sizeable damage or loss of lives. Differing opinions on possible exemptions were elicited, as most say that it must be employed objectively across all violators, although doctors should be given special consideration, especially when attending to urgent calls. Most thought that hands-free devices, like earphones, Bluetooth, and voice-to-text/call apparatus were safer methods of using handheld phones while driving because these devices did not impair drivers cognitively. They believed that most can get away with traffic violations, and using darkly tinted cars was one way of evading traffic enforcers. There was a pervading perception that laws in the country were generally poorly implemented and less exacting.

*3.1.8 Recommended countermeasures against distracted driving mainly due to cell phone use: S*uggested strategies include school-based measures, advertisements, and driving license regulation and technology. Education is required to instill public and road safety at an early stage. It is also essential to influence more senior and future road users. Another strategy is through advertising. The use of quadruple media, e.g., broadcasting, print, radio and social media, in depicting the dangers resulting from distracted driving may help road users realize its seriousness. This could, in turn, influence them to practice safe driving. The social networking sites were cited as major portals that could efficiently reach target audiences, such as the gadget-using millennials and novice drivers, through the use of viral videos of the catastrophic effects of risky driving practices. The "Facebook (FB) psychology" could be one measure to have a major effect on internet users. Viral videos on FB showing the 'drama' of the disastrous consequences of distracted driving could prove useful in conveying road safety messages. However, for drivers who may not have access to the internet or are not social media savvy, such as public or modified transport drivers (e.g., jeepneys, pedicabs, tricycles, etc.), signage on areas where they converge, such as eateries or jeepney terminals, should be placed warning them of the penalties and potential post-crash scenarios.

Stringent screening of prospective drivers should be enforced when securing driver's licenses, ensuring full recognition of road safety policies and traffic regulations. One proposed strategy is to increase road safety awareness by obligatory attendance to tailored lectures conducted by mandated traffic regulatory agencies e.g. the land transportation office, during periods of license renewal. Technology can also be used to catch violators of traffic laws. High definition cameras that can penetrate heavily tinted vehicles should be positioned in strategic places. A unique futuristic proposal entailing car engineering is to allocate a spot for telecommunications devices on the driver's side. When a gadget or phone is placed in this specially allocated carrier, it will automatically close, making access impossible while the car engine is running.

4. Discussion

The digital age has drastically escalated the possession and use of electronic gadgets. Farmer et al. in 2010 reports that drivers under 30 years old are likely to be distracted 16% of the time while driving [19]. Current investigation shows no significant difference in overall risk perceptions among drivers using a mobile phone from those who do not. Interestingly, a significant association has been noted concerning the perception of drivers using a cell phone: that hands-free devices are safer to use when driving. Compelling research evidence indicates that conversations on cell phones whilst driving, whether handheld or hands-free, increase the risk of injury and property damage crashes fourfold [20], [5]. Many drivers mistakenly consider talking on a hands-free cell phone safer than on a handheld phone [21]. These devices are erroneously seen as the safer solution to the risks of distracted driving because they help remove two apparent risks – the visual, looking away from the road, and the manual, taking one's hands off of the steering wheel. However, hands-free devices do not eliminate cognitive distraction, which can occur when the driver veers his mind off the road. Distracted drivers experience what researchers call inattention blindness, which has been compared to tunnel vision. They may be looking through the windshield, but their brains fail to process everything in the roadway environment that is necessary to monitor their surroundings sufficiently, identify potential hazards and respond to unexpected dangers on the road [22]. Using hands-free phones is more likely to cause drivers to miss relevant objects both in high and low places; this will certainly render them incapable of paying attention to more critical road details [23].

Most individuals recognize when they are visually or mechanically distracted and will usually disengage from these activities once they are fully aware. On the other hand, people are ordinarily unaware when distracted cognitively, such as conversing on the phone, resulting in increased in risk exposure time. Added to the dangers of hands-free phone use are the findings that this led to: an increase in reaction time in braking vehicles [24], to missing visual cues critical to safe navigation [25] and to lowered performance in safety tasks, such as peripheral visual checking and monitoring visual instruments such as the rear view and side mirrors [26].

The staggering prevalence of cell phone use of 90.68% in this study only reinforces the widely reported use of mobile phones while driving. An observational study undertaken in Australia confirms that young drivers (under 30 years) use mobile phones more often than middle-aged and older drivers (over 30 years old) while driving [27]. Driver distraction has already joined the ranks of alcohol and speeding as leading causes of fatal and serious road injury crashes. In 2010 the National Safety Council approximates that around 21% of all RTCs involved talking on cell phones, accounting for 1.1 million crashes the same year [28]. The results of the present study indicate that age, gender, annual family income, involvement in RTC and knowledge of anti-DD laws are not correlated with cell phone use; only driving experience of more than two years is significantly associated.

Overall, unsafe attitudes are significantly higher among drivers who use handsets while driving compared to those who do not ($p=0.007$), and the same significant result is also seen on the use of handsets while driving, even when it is known to be dangerous ($p=0.003$). Univariate logistic regression showed that driving experience of up to two years ($p=0.001$), risk perception ($p=0.046$), and attitude ($p=0.005$) as possible predictors, but further analysis using Multiple Logistic Regression revealed years of driving and attitude as the only significant predictors. These results are consistent with the conclusion of the European Survey on Road Users' Safety Attitudes (ESRA) 2015, indicating attitude as having a substantial effect on the self-declared behavior of sending text messages and emails while driving [29]. The recent study by Oviedo-Trespalacios on the risk factors of mobile phone use while driving in Queensland also identifies attitude as a predictor of cell phone engagement. On the other hand, some research indicates that novice, inexperienced drivers are more likely to engage in DD [30] and most describe younger drivers to be particularly prone to distraction [31]. These results, however, are not replicated in this study. The educational background and medical occupation of the present study participants could have an effect on the number of years they have been driving. The majority have mid to higher-range family incomes and jobs requiring mobility; most would have the capacity to drive for longer compared to similarly aged individuals from the general public, thus explaining the high proportion (87.59 %) of longer driving experience coupled to the high self-reported cell phone use while driving.

To understand distracted driving in young adults, a behavioral modification framework that can explain such risky behavior is the Theory of Planned Behavior (TBP). According to this model, intention is the most proximal determinant of behaviour, which is in turn influenced by attitude, subjective norm and perceived behavioral control (PBC). Attitude reflects an individual's favorable or unfavorable evaluation of performing a particular behavior; subjective norm refers to the social pressure a person feels in performing or refraining from a behavior and PBC pertains to self-efficacy or the degree to which one feels in engaging in a particular behavior [32]. The TBP constructs have satisfactorily explained, as a theoretical framework, the high level of mobile phone use among drivers [33].

The respondents in this research keep a very tight and demanding schedule, being resident trainees in a busy tertiary hospital. The high prevalence of cell phone use while driving can be explained by their intention to remain connected to their peers, co-workers, patients and hospital superiors. This is shown by the overall unsafe attitude of using cell phones, despite awareness of its dangers and an unsafe risk perception of using hands-free devices to meet the need to continually communicate, even when driving. The presence of ‘in-laws’ or significant others are the scenarios described in the FGD that will cause them to avoid distraction when driving, and which may represent the subjective norm they yield to. Their perceived behavioral control can be influenced by their medical education and training, family income, driving history, past involvement in RTC or previous experience with traffic law enforcers. Only driving experience of more than two years and general attitude towards cell phone use when driving are found to be significant predictors.

The proposed measure for countering this risky driving behavior is a more strict enforcement of laws regarding distracted driving, as attitude is one of the significant predictors of cell phone use while driving. One of the novel suggestions for reducing mobile phone use is through the use of technology. A futuristic innovation in car engineering of allocating a specialized gadget carrier inside the vehicle that will automatically limit access while the engine is running is well worth considering. This could reduce mobile phone and other electronic gadget usage while the vehicle is still moving. Another recommended technological approach is to develop a handset with a built-in ‘driving mode’ similar to the ‘flight mode’ integrated into most smartphones. This new mode will, once in use, disable all texting and answering functions of the phone, and will also automatically send a message or a signal to any caller that the receiver on the other end is driving and unavailable.

These distractions while driving are fast becoming ubiquitous and socially acceptable, turning such behavior into a social norm. The “reflex to answer a cell phone the moment it rings", mentioned in the FGD, can be explained by the feeling of many young adults of the need to ‘stay connected’, which in turn has influenced the routines in their daily lives, including driving practices. The habit of young drivers of checking their phones remains a major challenge for road safety authorities [34]. They no longer differentiate the setting they are in; these drivers answer calls or messages instinctively. This is certainly evident in this study, with a very high prevalence of cell phone use while driving; that is also echoed in the FGD. A possible behavioral modification approach is to place thematic advertisements in quadruple media identifiable to the target audience. Identification is an important element of testimonials, and can be used in commercials, as it relies on an individual connecting at a deep emotional level with a ‘character,' and with suggestions that lead to positive behavioral change [35]. By combining the strategies of identification and placing a social stigma on texting and cell phone use while driving, similar to other risky driving practices such as drunk driving or speeding, a stronger message to offending drivers will be conveyed. This approach intends to shift the social acceptance of driver distraction towards rejection, increasing the awareness of the ill-effects of handset use while driving, and thus increasing safe driving performance.

A possible limitation of this research is in under-reporting of risky distracted driving, as it is seen to be unlawful and socially undesirable despite employing anonymity in data gathering. The use of the self-reported questionnaire to determine the level of cell phone use while driving may not be fully reliable in measuring actual use and practice. The survey is also limited to the driving population of resident trainees in a tertiary hospital situated in an area considered to have high-density traffic. The generalizability of the results is thus restricted, and caution in applying them to other drivers in other parts of the country, or to the general population, is advised. Future research may perhaps consider other groups of drivers in a different environment in order to overcome sampling population constraints.

5. Conclusion

The high prevalence of cell phone use (texting, reading a text, calling or receiving calls) in the present study provides support for the findings of most researchers on this form of distracted driving. Although there was no significant difference in the overall risk perception among those using a mobile phone from those who do not, a significant association was noted on the perception that hands-free devices are safer to use when driving. This risk perception is considered unsafe by most studies. Overall, an unsafe attitude is higher among drivers operating mobile phones while driving, and the same significant result is seen on the unsafe attitude of using handsets, even when drivers are aware of its dangers. Driving experience of more than two years and attitude are the only significant predictors. Recommended countermeasures to such risky driving behavior include placing a social stigma on distracted driving through quadruple media advertisements, innovations in car engineering, the development of built-in telecommunications hardware and, lastly, a more strict and consistent enforcement of traffic laws.

6. References:

1. **WHO. Road traffic injuries-Fact Sheet. Geneva,** Switzerland. http://www.who.int/mediacentre/factsheets/fs358/en/ (2017). Accessed 30 Sept 2017
2. CDC. Distracted Driving. Retrieved from Centers for Disease Control and Prevention. https://www.cdc.gov/motorvehiclesafety/distracted_driving/index.html. (2016) Accessed 27 March 2017
3. NHTSA. Distracted Driving: facts and statistics. National Highway Traffic Safety Administration. https://www.distraction.gov/stats-research-laws/facts-and-statictic.html (2014) Accessed 20 February 2017
4. WHO. Mobile phone use: a growing problem of driver distraction. Geneva, Switzerland. http://www.who.int/violence_injury_prevention/publications/road_traffic/distracted_driving/en/index.html (2011). Accessed 30 Sept 2017
5. McEvoy S, Stevenson M, McCartt A, et al. (2005). Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study. *BMJ. 2005;331*:7514
6. Strayer D, Drews F, Crouch D. A comparison of the cell phone driver and the drunk driver. *Human Factors. 2006;48:*382-391.
7. Horey W, Wickens C. Examining the impact of cell phone conversations on driving using meta-analytic techniques. 48 (1), 196–205. *Human Factors. 2006;48*:1 p. 196-205.
8. Hendrick J, Switzer J. Hands-free versus hand-held cell phone conversation on a braking response by young drivers. *Perceptual and Motor Skills 2007;105*:2 p. 514-522.
9. NHTSA. Distracted Driving: facts and statistics. National Highway Traffic Safety Administration. https://www.distraction.gov/stats-research-laws/facts-and-statictic.html (2014) Accessed 27 Feb 2017
10. Sullman M. An observational study of driver distraction in England. *Transportation Research Part F 2012;15* p. 272-278.
11. PCHRD. Message sent: Cell phone distractions can kill. http://www.pchrd.dost.gov.ph/index.php/news/library-health-news/1173-message-sent-cell-phone-distractions-can-kill
12. Nurullah A, Thomas J, Vakilian F. The Prevalence of Cell Phone Use While Driving in a Canadian Province. Transportation Research Part F: Traffic Psychology and Behaviour 2013;*19*. p. 55-62.
13. RA 10913 Senate of the Republic of the Philippines. Republic Act 10913. Anti-Distracted Driving Act. https://www.senate.gov.ph/ republic_ acts/ ra%2010913.pdf (2015) Accessed 27 February 2017
14. Harrison M. College student's prevalence and perceptions of text messaging while driving. *Accident Analysis and Prevention 2011;43* p. 1516-1520.
15. Bergmark R, Giklich E, Guo R, Gliklich R. Texting while driving: The development and validation driving survey of the distracted driving survey and risk score among young adults. *Injury Epidemiology 2016;3*:7.
16. Womack K. *Understanding and addressing Distracted Driving in Texas.* Texas: United Services Automobile Association 2013.
17. National Highway Traffic Safety Administration. (2011). Traffic Safety Facts: Driver Electronic Device Use in 2010. Retrieved from http://www.distraction.gov/download/research-pdf/8052_TSF_RN_DriverElectronicDeviceUse_1206111_v4_tag.pdf
18. Governor's Highway Safety Association. Curbing distracted driving. http://www.ghsa.org/html/publications/survey/distraction2010.html. (2010) Accessed 15 Mar 2017
19. Farmer C, Braitman K, Lund A. Cell phone use while driving and attributable crash risk. *Traffic Inj Prev 2010;11* p. 466-470.

20. Redelmeier D, Tibshirani R. Association between cellular-telephone calls and motor vehicle collisions. *New England Journal of Medicine 1997;336*:7 p. 453-458.
21. AAA Foundation. Cell phones and driving: research update. Retrieved from https://www.aaafoundation.org/sites/default/files/CellPhonesandDrivingReport.pdf. (2008) Accessed 15 March 2017
22. Maples W, DeRosier W, Hoenes R, Bendure R, Moore S. The effects of cell phone use on peripheral vision. Optometry. *Journal of the American Optometric Association 2008;79*:1 p. 36-42.
23. Strayer D, Cooper J, Drews F. What Do Drivers Fail to See When Conversing on a Cell Phone? (2014)
24. Strayer D, Drew F, Johnston W. Cell-phone induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied 2003;9*:1 p. 23-32.
25. Drews F, Pasupathi M, Strayer D. Passenger and cell phone conversations in simulated driving. 14. *Journal of Experimental Psychology: Applied 2008;14* p. 392–400.
26. Harbluk J, Noy Y, Trbovich P, Eizenman M. An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance. *Accident Analysis and Prevention 2007;39*:2 p. 372-378.
27. Young K, Rudin-Brown C, Lenne M. Look Who's Talking! A Roadside Survey of Drivers' Cell Phone Use. *Traffic Injury Prevention 2010; 11*:6 p. 555-60.
28. Kolosh K. *Summary of Estimate Model*. National Safety Council. http://www.nsc.org/news_resources/Resources/Documents/NSC Estimate Summary.pdf (2012) Accessed 30 September 2017
29. Areal A. Texting distracted driving behaviour among European drivers: influence of attitudes, subjective norms and risk perception. *Prevencao Rodoviaria Portuguesa.* Paris, France. www.esranet.eu (2017) Accessed 15 March 2017
30. Oviedo-Trespalacios O, King M, Haque M, Washington S. Risk factors of mobile phone use while driving in Queensland: Prevalence,attitudes, crash risk perception, and task-management strategies. *PLoS ONE 2017;12*:9 e0183361.
31. Hosking, S, Young K, Regan M. The effects of text messaging on young drivers. *Human Factors 2009;51*:4p. 582-592.
32. Ajzen I. The theory of planned behavior. *Organizational Behavior and Human Decision Processes 1991;50*:2 p. 179-211.
33. Walsh S, White K. Me, my mobile and I: The role of self and prototypical identity influences in the prediction of mobile phone behavior. *Journal of Applied Social Psychology 2007;37* p. 2405-2434.
34. Turnbull N, Algie J. A qualitative analysis of young driver's perceptions of driver distraction social marketing interventions. *World Social Marketing Conference*, (pp. 1-7). Sydney, Australia. http://ro.uow.edu.au/buspapers/813 (2017) Accessed 15 March 2017
35. Hinyard L, Kreuter M. Using Narrative Communication as a Tool for Health Behaviour Change: A Conceptual, Theoretical, and Empirical Overview. *Health Education and Behaviour 2007; 34*:5 p. 777-792.

7. Appendices

Table 6. Univariate Logistic Regression Analysis Predicting Cell Phone Use while Driving

	Odds ratio (OR)	95% CI	p- value p<0.05
1 Age	1.23	0.82-1.84	0.324
2. Gender (male=ref)			
female	0.38	0.12-1.17	0.091
3. Annual Family Income			
P100,000 and less (<USD1,935)= ref			
P100,001 to P500,000 (USD1,935- 9,679)	0.53	0.06-4.90	0.576
P500,001 to P1,000,000 (USD-9,679- 19,357)	0.86	0.09-8.54	0.900
P1,000,001 and above (>USD 19,357)	1.70	0.16-17.86	0.660
4. Involvement in RTC? (No=refe)			
Yes	1.30	0.45-3.77	0.633
5. Years of driving			
$\leq$ 2 years (ref)			
> 2 years	7.06	2.22-22.46	**0.001**
6. Knowledge about Anti Distracted Driving Law (No=ref)			
Yes	3.054	0.57-16.25	0.190
7. Risk Perception (Safe Risk Perception= Ref)			
Unsafe Risk Perception	3.66	1.02-13.16	**0.046**
8. Attitude on Using Cell phone while Driving (Safe Attitude= Reference)			
Unsafe Attitude	4.93	1.62-14.98	**0.005**

Table 7. Summary of Multiple Logistic Regression Predicting Use of Cellphone while Driving

Variable	Full Model (All Predictors)	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	OR (p-value)	OR (p-value)	OR (p-value)	OR (p-value)	OR (p-value)	OR (p-value)	OR (p-value)
Age	1.00 (0.987)	-	-	-	-	-	-
Gender							
Male (Reference)	-	-	-	-	-	-	-
Female	0.74 (0.657)	0.74 (0.656)	0.70 (0.576)	0.68 (0.543)	-	-	-
Annual Family Income							
P100,000 and less (<USD1,935)= Ref	-	-	-	-	-	-	-
P100,001 to P500,000 (USD1,935- 9,679)	0.94 (0.959)	0.94 (0.959)	-	-	-	-	-
P500,001 to P1,000,000 (USD-9,679-19,357)	2.85 (0.431)	2.85 (0.431)	-	-	-	-	-
P1,000,001 and above (>USD 19,357)	2.26 (0.526)	2.26 (0.523)	-	-	-	-	-
Involvement in RTC							
No (Reference)	-	-	-	-	-	-	-
Yes	1.43 (0.597)	1.43 (0.597)	1.11 (0.870)	-	-	-	-
Knowledge about Anti-Distracted Driving Law							
No (Reference)	-	-	-	-	-	-	-
Yes	4.77 (0.181)	4.80 (0.156)	3.26 (0.240)	3.09 (0.250)	3.17 (0.240)	4.01 (0.135)	-
Risk Perception							
Safe Risk (Ref)	-	-	-	-	-	-	-
Unsafe Risk	1.33 (0.765)	1.32 (0.763)	2.30 (0.284)	2.34 (0.273)	2.42 (0.252)	-	-
Attitude							
Safe Attitude (Ref)	-	-	-	-	-	-	-
Unsafe Attitude	2.91 (0.147)	2.92 (0.141)	3.20 (0.081)	3.18 (0.082)	3.27 (0.072)	**4.01 (0.028)**	**3.61 (0.039)**
Years of Driving							
$\leq$2 years (Reference)	-	-	-	-	-	-	-
> 2 years	**6.52 (0.009)**	**6.50 (0.006)**	**5.53 (0.007)**	**5.58 (0.007)**	**6.12 (0.003)**	**6.32 (0.003)**	**6.35 (0.002)**