<table>
<thead>
<tr>
<th>Attitudes and preferences</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Discomfort detection in automated driving by psychophysiological parameters from smartbands</td>
<td><strong>Matthias Beggiato, Franziska Hartwich, Josef Krems – TUC</strong></td>
</tr>
<tr>
<td>• Designing the transition from highly automated driving to manual driving: a study of driver’s preferences</td>
<td><strong>Stefan Brandenburg, Sandra Epple - TU Berlin</strong></td>
</tr>
<tr>
<td>• Evaluation of free public transport for older people in Sweden</td>
<td>**Tania Wilstrand, Per Henriksson, Lena Levin VTI</td>
</tr>
<tr>
<td>Distraction and innattention – Parallel session</td>
<td></td>
</tr>
<tr>
<td>• Self-regulation of drivers’ mobile phone use: the influence of driving context</td>
<td><strong>Michiel Christoph, Simone Wesseling, Nicole van Nes - SWOV</strong></td>
</tr>
<tr>
<td>• What are drivers doing when they are not on the cell phone?</td>
<td>**Carol Flannagan - University of Michigan</td>
</tr>
<tr>
<td>• Effects of secondary tasks and display position on glance behavior during partially automated driving</td>
<td>**Ann Christin Hensch, C. Schmidt, N. Rauh, J. Krems - TUC</td>
</tr>
<tr>
<td>• ESRA: E-Survey of road user’s safety attitudes - Analysis of safety indicators and predictors of distracted driving behaviours</td>
<td>**Veerle Ross, Kris Brijs, Muhammad Wisal Khattak - IMOB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Automated vehicles: trust, accept and move on? – Parallel session</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Fostering trust and acceptance of a collision avoidance system through retrospective feedback</td>
<td><strong>David Large, James Khan, Gary Burnett - University of Nottingham</strong></td>
</tr>
<tr>
<td>• Measuring trust in automated vehicles through gaze bahaviour</td>
<td><strong>Francesco Walker, Willem Verwey, Marieke Martens - University of Twente</strong></td>
</tr>
<tr>
<td>• Understanding trust in an AD-context: a mixed method approach</td>
<td><strong>Fredrick Ekman, Mikael Johansson, MariAnne Karlsson - Chalmers</strong></td>
</tr>
</tbody>
</table>
• Vehicle movements as implicit communication signal between pedestrians and automated vehicles

_Claudia Ackermann, Matthias Beggiato, Luka-Franziska Bluhm, Josef Krems – TUC_

Driver behavior and training

• Education of future car drivers in Flanders

_Roel de Klerk, Kim Jacobs - VSV_

• Automated feedback on viewing skills lowers accident involvement

_Jorrit Kuipers, Peter Wieringa - TU Delft_

• License to Supervise. Influence of Driving Automation on Driver Licensing

_Arie Van den Beukel - University of Twente_

Driver state and workload

• The assessment of hazard awareness skills among light rail drivers

_Avinoam Borowsky - Ben Gurion University_

• Discriminating drivers’ fear and frustration through the dimension of power

_Meng Zhang - DLR_

• Review of medical fitness to drive in Europe

_Brian Fildes, Andrew Morris - Loughborough University | Jennie Oxley - Monash University Accident Research Centre | Shaun Helman, Jill Weekley - TRL Limited_

• Driver’s recovery performance in a critical run-off-road scenario - A driving study -

_Alexander Eriksson, Bruno Augusto, Niklas Strand, Jesper Sandin – VTI_

Poster pitches

• The organizational response to automation support degradation. Identifying air traffic control sources of resilience in cases of radar loss

_Luca Save, Daniele Ruscio - Deep Blue | Valentina Cedrini, Maurizio Mancini - ENAV_

• Human centred design recommendations for autonomous car in transition phases

_Annie Pauzie, Ferhat Lyess - IFSTTAR_

• Meaningful human control over automated driving systems

_Daniel Heikoop, Marjan Hagenzieker, Giulio Mecacci, Filippo Santoni de Sio, Simeon Calvert, Bart van Arem - TU Delft_
Consumers’ perceptions towards Autonomous and Connected Vehicles: a Focus-Group Survey on University Population

*Ilias Panagiotopoulos, George Dimitrakopoulos - Harakopio University of Athens*

Measuring driver behavior and the influence of context – Parallel session

- Heart rate analysis for human factors: Development and validation of an open source toolkit for noisy naturalistic heart rate data
  
  *Paul van Gent, Haneen Farah, Bart van Arem - TU Delft | Nicole Van Nes - SWOV*

- Harsh braking by truck drivers: a comparison of thresholds and driving contexts and driving contexts using naturalistic driving data
  
  *Reinier Jansen, Simone Wesseling - SWOV*

- A normal driving based deceleration behaviour study towards autonomous vehicles
  
  *Stravoula Panagiota Deligianni, Mohammed Quddus, Andrew Morris, Aaron Anvuur - Loughborough University*

- Investigation of Herringbone pattern and Optical Circles for Safe Driving Behaviour at Curves Using Driving Simulator
  
  *Hammad Hussain Awan, Ali Pridavani, Tom Brijs – IMOB*

HMI design, how to get it right? – Parallel session

- Augmented reality as an ADAS system: a cognitive approach
  
  *Lucas Morillo, Eva Maria Garcia Quinteiro - CTAG*

- Ipsilateral versus contralateral tactile alerts for take-over requests in highly automated driving
  
  *Guy Cohen-Lazry, Nuphar Katzman, Tal Oron-Gilad - Ben Gurion University*

- Lane change manoeuvres for automated freeway driving applications
  
  *Klas Christoph, Nicole Eikenlenberg, Gudrum Voss, Louis Tijerina, Peter Zegelaar - RWTH Aachen University | Avinoam Borowsky - Ben Gurion University*

- "Trust me - I’m AutoCAB": using natural language interfaces to improve the trust and acceptance of level 4/5 autonomous vehicles
  
  *Vicki Antrobus, Gary Burnett - University of Nottingham*

- A 'driver-more' Approach to Vehicle Automation
  
  *Andrea Catallano, Serena Fruttaldo, Elisa Landini, Roberto Montanari - RE:Lab | Andreas Luedtke - Offis*
ATTITUDES AND PREFERENCES
Discomfort Detection in Automated Driving by Psychophysiological Parameters from Smartbands

Matthias Beggiato, Chemnitz University of Technology, Germany, matthias.beggiato@psychologie.tu-chemnitz.de, Franziska Hartwich, Chemnitz University of Technology, Germany, Josef Krems, Chemnitz University of Technology, Germany

ABSTRACT

The research project KomfoPilot at Chemnitz University of Technology aims at assessing discomfort in automated driving using psychophysiological parameters from smartbands. In an empirical driving simulator study, 40 participants from 25 to 84 years experienced two highly automated drives including three potentially critical and discomfort-inducing approach situations in each drive. The own car drove in automated mode with 100 km/h and approached a truck driving ahead at a constant speed of 80 km/h. Automated braking started very late at a distance of 9 m reaching a minimum distance of 4.2 m and minimum time to contact (TTC) of 1.1 s. Participants reported perceived discomfort continuously by a handset control integrated into the driving simulator (Hartwich et al., 2015, 2018). Psychophysiological parameters were assessed using the Microsoft Band 2 and included heart rate (HR), heart rate variability (HRV) and skin conductance level (SCL). To analyse the potential of band data for discomfort detection, psychophysical metrics during discomfort periods were compared to the values at 10 s time intervals prior and after. HR decreased during discomfort periods, HRV showed the expected u-shaped pattern with a decrease during the discomfort intervals, and after correcting for linear growing trend, SCL decreased as well. Overall, psychophysiological metrics showed potential to detect discomfort and will therefore be included in the detection algorithm. One of the challenges for using smartbands will be the use of adequate signal analysis methods for gaining the maximum signal-to-noise ratio.

Keywords: discomfort, automated driving, smartbands, psychophysiological parameters.

1 BACKGROUND AND OBJECTIVES

Wearable devices such as smartbands / fitness trackers gain increasing popularity and offer cheap and easy-to-use assessment of psychophysiological parameters in daily live situations such as driving. With increasing vehicle automation, smartbands could provide valuable information about driver states such as discomfort to improve human-machine collaboration. Detected discomfort could subsequently be used to adapt driving parameters as well as information presentation. Comfort is understood as a subjective, pleasant state of relaxation given by confidence and an apparently safe vehicle operation (Constantin, Nagi, & Mazilescu, 2014), “which is achieved by the removal or absence of uneasiness and distress” (Bellem et al., 2016, p. 45). Next to safety and efficiency, the potential to increase driving comfort is considered one of the main motivations for forwarding driving automation (European Road Transport Research Advisory Council 2017). The research project KomfoPilot at Chemnitz University of Technology aims at the assessment of discomfort in automated driving using psychophysiological parameters from smartbands as one data source. Overall goal of the project is the development of an algorithm for real-time discomfort detection to subsequently adapt driving style and information presentation in real-time. The use of commercially available smartbands is an explicit project goal to estimate the potential and challenges of such devices. The present paper reports the results of the psychophysiological metrics Heart Rate (HR), Heart Rate Variability (HRV) and Electrodermal Activity (EDA) with regard to perceived discomfort during automated driving. All metrics were assessed in a driving simulator study using the smartband Microsoft Band 2 (Details on the MS Band 2 in Schmalfuß
et al., 2018). The use of these metrics to infer mental states such as stress, workload, arousal, fear, panic... etc. has a long research tradition (overviews in Cowley et al., 2015; Backs & Boucsein, 2000; Andreassi, 2000; Schandry, 1998). Study results are not uniform, however, overall tendencies can be transferred into the topic of discomfort detection. HR and Skin Conductance Level (SCL) usually increases with physical, mental and emotional strain, whereas HRV decreases. Thus, we expected an increase in HR and SCL during discomfort periods with a return to the baseline levels after these situations (inverse U-shaped pattern) and the opposite for HRV (U-shaped pattern).

2 METHOD

Study design: The study was conducted in a fixed-base driving simulator (Software SILAB 5.1) with a fully equipped interior and a 180° horizontal field of view, including a rear-view mirror and two side mirrors. All 40 participants took part in two distinct driving sessions with approximately two months delay in between. In each of the two sessions, all participants experienced an identical 3 minutes highly automated drive including three potentially critical and discomfort-inducing approach situations. The own car drove in automated mode with 100 km/h and approached a truck driving ahead at a constant speed of 80 km/h. Automated braking started very late at a distance of 9 m reaching a minimum distance of 4.2 m and minimum time to contact (TTC) of 1.1 s (Fig. 1 left). Perceived discomfort was assessed continuously by a handset control integrated into the driving simulator (Hartwich et al., 2015, 2018; Fig. 1 right). Participants were instructed to press the lever in accordance with the extent of perceived discomfort and had no possibilities to intervene using pedals or steering wheel.

Participants: The sample consisted of 40 participants (25 male, 15 female) ranging from 25 to 84 years. The younger group (25 to 45 years) included 21 persons with a mean age of 30 years (SD = 4.3). A total of 19 persons formed the older group (65+ years) with a mean age of 72 years (SD = 6.0). None of the persons had previously experienced automated driving in the driving simulator. Participants signed an informed consent and were compensated with 20 Euro for each session.

Sensors and interval selection: Psychophysiological parameters were assessed continuously using the Microsoft Band 2 (Fig. 1 right) and included HR, HRV and SCL. In addition, accelerometer and gyroscope data was recorded from the band sensors to correct for movements. The MS Band 2 comes with a Software Development Kit, which allowed for programming a dedicated logging application via Bluetooth connection. The complete study also comprised additional sensors which are not part of these analyses such as Eye-Tracking (SMI Eye Tracking Glasses 2), marker-based Motion Tracking (OptiTrack), a seat pressure mat, two 3D-cameras and six 2D-cameras. To analyse the potential of band data for discomfort detection, psychophysiological metrics during discomfort periods were compared to the values at 10 s time intervals prior and after (Fig. 2).
Discomfort periods were extracted from the beginning of pressing the handset control lever until the lever was released – independent of the magnitude. The 40 participants experienced 6 approach situations in total, which would result in 240 situations. However, the handset control was only pressed in 208 situations. In addition, single data channels from the band were not recorded in some situations due to technical reasons. Finally, 206 discomfort periods entered the analysis for HR (M = 8.10 s, SD = 5.52 s), 202 sequences for HRV (M = 8.10 s, SD = 5.53 s) and 203 sequences for EDA (M = 8.16 s, SD = 5.51 s). Data preparation procedures are described in the results section for each sensor channel.

3 RESULTS

Heart rate: Raw HR-values in beats per minute were recorded with 1 Hz frequency from the MS Band 2. To correct for interindividual variability (Jennings & Allen, 2017), the raw values were transformed into z-scores for each of the 206 discomfort sequences including the 10 s before and after. Fig. 3 reports the means of these z-scores over the 206 sequences (repeated measures ANOVA with Greenhouse-Geisser correction and Bonferroni-adjusted post-hoc tests). HR decreased significantly in the discomfort interval compared to the 10 s before, however, HR did not return to the previous level in the 10 s after. A detailed timeline-plot of the mean z-scores showed that there was indeed a return to the baseline level, but the increase started only about 5 s after the end of the discomfort interval. Basically, a u-shape was present, but delayed for approximately 5 s with regard to the discomfort interval.

---

**Figure 2** – Smartband data and three discomfort interval selections with 10 s interval before and after

**Figure 3** – Mean Heart Rate z-scores for discomfort intervals and 10 s before/after
Heart Rate Variability: To calculate HRV metrics, the inter-beat-intervals (IBI) values in seconds were recorded from the MS Band 2 with a new value for each detected IBI (no fixed frequency). In the specific case of the MS Band 2, HR and IBI are not reciprocal values, but IBI is recommended to be used for HRV-calculations (Cropley et al., 2017). The Root Mean Square of Successive Differences (RMSSD) was calculated for the discomfort interval and the 10 s prior and after. RMSSD is considered the best parameter for short periods and intervals with unequal length (Berntson et al., 2017). Mean RMSSD-scores over the 202 intervals showed the expected u-shaped pattern (Fig. 4) with a statistically significant decrease in HRV during the discomfort interval compared to the 10s prior and after ($\chi^2(2) = 40.05, p < .001$, Friedman’s non parametric ANOVA).

Electrodermal Activity: The MS Band 2 measured skin resistance in kilo ohm with a frequency of 5 Hz using two electrodes on the opposite site of the display (see Fig. 1 right). Raw values were inverted to obtain the skin conductance level (SCL) in micro Siemens. As the EDA values were very sensitive to hand movements (e.g. placing the hand on the knees), SCL values during high movement episodes were excluded on the basis of the Band accelerometer and gyroscope data. Similar to HR, raw SCL values were transformed into z-scores for each of the 203 discomfort sequences including the 10 s before and after. A detailed timeline-plot of the mean z-scores showed a linear continuous increase in SCL over time, independent of the situation. In order to correct for this general linear trend, a linear regression was calculated for each sequence and subtracted from the z-scores to get the detrended z-scores. The mean detrended z-scores are reported in Fig. 5. Contrary to the expectations, results showed a u-shaped pattern for EDA with a decrease of SCL during discomfort intervals, however, with a small effect size of $\eta^2_p = .025$. 

Figure 4 - Mean RMSSD (Heart Rate Variability) for discomfort intervals and 10 s before/after

Figure 5 - Mean z-score of detrended Skin Conductance Level for discomfort intervals and 10 s before/after
4 DISCUSSION

The driving simulator study within the research project KomfoPilot at Chemnitz University of Technology aimed at assessing discomfort in automated driving using psychophysiological parameters from the smartband MS Band 2. Overall, the psychophysiological parameters HR, HRV and EDA showed changes associated with the perceived discomfort indicated by the handset control. In contrast to the hypothesis, HR decreased during discomfort periods. A possible explanation for this phenomenon could be the effect of “preparation for action”, which means an anticipatory deceleration of HR prior to actions (Cooke et al. 2014; Schandry, 1998). This effect was reported for sport actions, but also for simpler reaction time paradigms; “it is well established that HR deceleration occurs during the fixed foreperiod of an RT task” (Andreassi, 2000, p. 270). HRV showed the expected u-shaped pattern with a decrease during the discomfort intervals. Raw EDA showed a linear increasing trend over time, which could be explained by the fact that participants got warm during driving. After correcting for this linear trend, EDA showed a slight decrease during reported discomfort, which is contrary to the expected evolvement. However, the effect size was small and the inverse effect could be related to measurement procedures associated with the smartband: Firstly, absolute EDA values were highly dependent on how tight the band was closed. These differences could be corrected using z-scores, however, some bias could remain e.g. when the band was worn very loosely. Secondly, EDA measures were taken from the outer side of the wrist, which is a much less sensitive place for SCL-changes compared to e.g. the fingers (Andreassi, 2000). Thirdly, hand movements caused partly strong effects/offsets in EDA values. The applied quite simple correction method of excluding these parts could potentially be improved by more sophisticated algorithms such as e.g. forward prediction and offset correction. The mentioned problems such as e.g. less control on how tight the band is closed are to some extend related to the use of smartbands instead of more sophisticated measurement devices. However, the aim of the project was and is to assess the potential of existing wearable devices with all the real-world usage challenges. Even with these problems, effects related to discomfort could still be identified in the data. One of the major challenges for using these devices to detect discomfort will be the use of adequate signal analysis methods for gaining the maximum signal-to-noise ratio. Additional improvements in detection could be achieved by the joint/multivariate analysis of these psychophysiological parameters including additional metrics such as eye-tracking, body movements, vehicle kinematic and situation information. These analyses and the development of a data fusion algorithm are the next steps in the project.

ACKNOWLEDGEMENT


REFERENCES


Designing the Transition from Highly Automated Driving to Manual Driving: A Study of Drivers’ Preferences

Stefan Brandenburg, Technische Universität Berlin, Germany, stefan.brandenburg@tu-berlin.de,  
Sandra Epple, Technische Universität Berlin, Germany

ABSTRACT

One important aspect of highly automated driving is the transition of control between automated vehicle and driver. Previous research showed that the design of take-over requests (TORs) influences the success of this shift of control. In an online study with 53 participants (26 females), we examined drivers’ preferences regarding TOR modality (auditory and visual) and TOR procedure (one or two steps). In one-step TOR procedures a TOR is presented at a single point in time. In contrast, in two-step TOR procedures drivers are informed about the transition of control at two points in time: an initial warning followed by an alarm. The study’s findings show that a two-step TOR procedure is preferred to a one-step TOR procedure. Two-step TORs are rated as more intuitive, useful, attractive and more appropriate in displaying information than one-step procedures. Moreover, participants preferred verbal auditory TORs (speech) to non-verbal TORs (tone). Implications on the design of TOR interfaces for highly automated driving are discussed.

Keywords: Highly automated driving, HMI design, take-over request, subjective preferences, TOR modality, two-step TOR procedure.

5 INTRODUCTION

Highly automated cars are estimated to be on European roads between 2020 and 2025 (ERTRAC, 2015). They were anticipated for a long time and became more likely through the advances in computerisation, the development of on-board sensors (Walker, Stanton, & Young, 2001) and changes in legal restrictions (ECE/TRANS/WP.1/145, 2014). Amendments made to the 1968 Vienna Convention on road traffic, allow for the transition of vehicle control to an automated system, as long as drivers can resume control whenever needed (ECE/TRANS/WP.1/145, 2014). The transition of control between driver and automated vehicle is an important and safety-critical aspect of highly automated driving (e.g. Merat, Jamson, Lai, Daly, & Carsten, 2014). In order to make this transition of control as safe and comfortable as possible, drivers have to be provided with an understanding of the driving situation and of the steps necessary to resume control of the vehicle in a safe and comfortable manner. The design of take-over requests (TOR) can accelerate drivers’ understanding of the situation and of necessary actions. The human-machine-interface (HMI) issues take-over requests and mediates the interaction between vehicle and driver. Therefore, the design of the HMI has a significant impact on safety outcomes of automated systems (Casner, Hutchins, & Norman, 2016). For instance, HMI design influences drivers’ reaction times when resuming control of the vehicle (Forster, Naujoks, Neukum, & Huestegge, 2017). Ideally, drivers are able to intuitively understand the interface that issues a take-over request because intuitive interaction is fast, unconscious and automatic (Macaranas, Antle, & Riecke, 2015). But how should take-over requests be designed? Is there a combination of HMI design aspects that results in take-over requests with high usability, usefulness, and attractiveness? In the past, various different HMI designs have been used to issue take-over requests to drivers. Design aspects that have been examined in this context include TOR modality (e.g. auditory and visual TORs) and TOR procedure (e.g. one-step and two-step procedures). The term TOR procedure refers to the
number of take-over requests that are presented within one take-over situation. In one-step TOR procedures a TOR is presented at a single point in time. In contrast, in two-step TOR procedures drivers are informed about the transition of control at two points in time: an initial warning followed by an alarm. Two-step TOR procedures have the potential to create a more gradual take-over process, by providing a time frame between warning and alarm, within which the driver can take-over control. According to Walch, Lange, Baumann, and Weber (2015), two-step TOR procedures are preferred to one-step TOR procedures. However, the authors did not investigate the impact of TOR modality on two-step TOR procedures. Research on TOR modalities has focused on auditory, visual, and tactile modalities in one-step TOR procedures (e.g. Forster et al., 2017; Politis, Brewster, & Pollick, 2015). The results of these studies suggest that verbal auditory TORs (speech) can be valuable for the design of TOR interfaces. According to Forster et al. (2017), adding verbal auditory information (speech) to non-verbal visual-auditory TORs lead to shorter reaction times and more positive subjective ratings. These findings are in contrast with many industrial prototypes that mostly rely on written text or pictograms appearing on the dashboard in combination with a single tone. Moreover, past research on TORs has focused on one-step TOR procedures only. Yet, no studies tested whether these findings are valid for two-step TOR procedures. The present study aims to close this gap by systematically varying TOR procedure and TOR modality. The first objective of this study is to examine the effect of TOR procedure (one-step or two-step) on drivers’ preferences. The second objective is to investigate the impact of TOR modalities (auditory and visual) on drivers’ preferences.

6 MATERIAL AND METHODS

Participants

A total of 53 participants (26 females) took part in the study. Their mean age was 32 years (SD = 9.86 years) and ranged from 20 to 64 years. Forty-eight participants (91%) possessed a valid drivers’ license. The study was subject to evaluation of the local ethics committee.

Design of the human-machine interfaces for take-over requests

Participants experienced 8 different human-machine interfaces issuing take-over requests. These TOR interfaces differed regarding three dimensions: (1) TOR procedure (one-step vs. two-step), (2) the presentation of auditory information (via tone vs. via speech), and (3) the presentation of visual information (via text vs. via text and pictogram). In the condition with a one-step TOR procedure, an alarm was issued at a single point in time (“Alert - Take-over vehicle control now”). A two-step TOR procedure, on the other hand, consisted of an initial warning (“Warning - Roadworks ahead - Take-over vehicle control soon”) followed by the alarm (“Alert - Take-over vehicle control now”). All TOR interfaces were multimodal as they contained auditory as well as visual information. Auditory information was either presented by a single tone, or by a mechanical voice reading the warning and the alarm (speech). Visual information was presented by a written text, or by text and additional pictograms appearing on the dashboard. The pictograms displayed a standard road works sign for the warning and the vehicles’ pedals and steering wheel for the alarm if applicable. TOR interfaces were dynamic in the sense that they were presented as short film clips. For the one-step TOR the film clip was 8 seconds long and for the two-step TOR 16 seconds long. Each film clip followed the same plot. First, the dashboard was displayed for 3 seconds. Then the TOR appeared and lasted for 3 seconds. Finally, the TOR faded and the dashboard screen was displayed again for 2 seconds. In the two-step TOR procedure, two clips were displayed: an initial warning followed by the alarm. Examples of the TOR interfaces used in this study can be found in table 1.
Table 1 – Screenshots of 2 examples of the take-over request interfaces

<table>
<thead>
<tr>
<th>Interface 1</th>
<th>Interface 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>one-step procedure with text and tone</td>
<td>two-step procedure with text, pictogram, and speech</td>
</tr>
<tr>
<td>![Interface 1 Screenshot]</td>
<td>![Interface 8 Screenshot]</td>
</tr>
</tbody>
</table>

Procedure
Participants were invited to this online-study via social media platforms like facebook™ and via university mailing lists. When they clicked on the invitation link they were forwarded to SociSurvey, the online tool that was used for data assessment. The online study consisted of four parts: (1) an introduction, (2) a general evaluation of each TOR interface, (3) a ranking of TOR interfaces, and (4) a detailed evaluation of the personal best and worst TOR interfaces. Part (4) will not be discussed any further in this paper due to limited space. First, participants read information about the general procedure and gave their informed consent and data assessment agreement. Then, a short introductory text instructed participants to imagine sitting in a highly automated car and reading text messages on their smart phone. After that, they experienced each of the 8 TOR interfaces separately and in randomized order. Participants were asked to evaluate each TOR interface regarding the following four items: (a) This TOR interface is intuitive, (b) I find this TOR interface useful, (c) I find this TOR interface attractive, and (d) I find the amount of information appropriate. Finally, participants sorted the interfaces in descending order with the best interface on rank 1 and the worst interface on rank 8.

7 RESULTS
a. General Evaluation of TOR interfaces
A within-subjects MANOVA revealed significant main effects for each of the independent variables: TOR procedure, visual modality, and auditory modality. Participants rated two-step TOR procedures as more intuitive $(F(1, 50) = 29.22, p < .001, \eta^2_{\text{part}} = .36)$, more useful $(F(1, 50) = 24.87, p < .001, \eta^2_{\text{part}} = .33)$, more attractive $(F(1, 50) = 21.35, p < .001, \eta^2_{\text{part}} = .29)$, and more appropriate for displaying the information $(F(1, 50) = 25.80, p < .001, \eta^2_{\text{part}} = .34)$ than one-step TOR procedures. Moreover, participants indicated that speech was more intuitive $(F(1, 50) = 5.14, p = .02, \eta^2_{\text{part}} = .09)$, more useful $(F(1, 50) = 5.54, p = .02, \eta^2_{\text{part}} = .10)$. Lastly, text and pictogram received higher ratings with respect to intuitiveness $(F(1, 50) = 6.38, p = .01, \eta^2_{\text{part}} = .11)$, and attractiveness of the TOR interface $(F(1, 50) = 7.86, p = .01, \eta^2_{\text{part}} = .13)$ than mere text. Moreover, the MANOVA revealed significant interactions between auditory and visual modality on usefulness $(F(1, 50) = 4.62, p = .03, \eta^2_{\text{part}} = .08)$ and appropriateness of information $(F(1, 50) = 4.56, p = .03, \eta^2_{\text{part}} = .08)$. Text and pictogram were rated as more useful and as more appropriate in displaying information than pure text, when a tone was used as auditory information. However, when speech was used as auditory information, there was no
difference between the two visual modality conditions with respect to usefulness and appropriateness of information. All other main effects and interactions were not significant.

b. TOR interface ranking

Analysing the ranking of the 8 different TOR interfaces, results showed that each interface appeared on all possible ranks at least once. However, clear preferences became evident when analysing the median rank of each TOR interface. Friedman’s ANOVA for non-parametric data showed a significant difference indicating that the rank distributions of the interfaces are not similar, p < .001. Table 2 shows each interface, its median rank and the corresponding interquartile range as measure of spread.

Table 2 – Median ranks of take-over request (TOR) interfaces

<table>
<thead>
<tr>
<th>Rank</th>
<th>Interface</th>
<th>Median rank (interquartile range)</th>
<th>Design aspect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>One-step or two-step</td>
<td>Text or text + pictogram</td>
</tr>
<tr>
<td>1 (best)</td>
<td>No. 7</td>
<td>2 (1-2.5)</td>
<td>Two-step</td>
</tr>
<tr>
<td>2</td>
<td>No. 8</td>
<td>3 (1-5)</td>
<td>Two-step</td>
</tr>
<tr>
<td>3</td>
<td>No. 5</td>
<td>4 (2-5)</td>
<td>Two-step</td>
</tr>
<tr>
<td>4</td>
<td>No. 6</td>
<td>4 (3-5)</td>
<td>Two-step</td>
</tr>
<tr>
<td>5</td>
<td>No. 3</td>
<td>5 (3.5-6)</td>
<td>One-step</td>
</tr>
<tr>
<td>6</td>
<td>No. 4</td>
<td>6 (5-6)</td>
<td>One-step</td>
</tr>
<tr>
<td>7</td>
<td>No. 1</td>
<td>6 (4-7)</td>
<td>One-step</td>
</tr>
<tr>
<td>8 (worst)</td>
<td>No. 2</td>
<td>7 (6-8)</td>
<td>One-step</td>
</tr>
</tbody>
</table>

Note. The interquartile range is a measure of spread that captures 50% of the values around the median.

Table 2 shows that TOR interface No. 7 received the lowest median rank and therefore was rated the best interface. It combined text with speech in a two-step procedure. Interface No. 8, which contained an additional pictogram, received the second lowest median rank. The TOR interface with the highest rank (worst interface) was interface No.2. It consisted of a single tone and a combination of text and pictogram in a single step. Moreover, table 2 shows that all two-step TOR conditions outranked all one-step TOR conditions. Within the two-step and one-step TOR procedure conditions, speech outranked pure tone.

8 DISCUSSION

The present study had two objectives. The first objective was to examine the effect of TOR procedure (one-step or two-step) on drivers’ preferences. The second objective was to investigate the impact of TOR modalities (auditory and visual) on drivers’ preferences.

Regarding the first objective, our findings show that the two-step TOR procedure is preferred to the one-step TOR procedure. The two-step procedure was rated as more intuitive, useful, attractive and more appropriate in displaying the relevant information than the one-step procedure. Moreover, median ranks indicate that all two-step TOR conditions outrank all one-step TOR conditions. In line with the study of Walch et al. (2015), our findings suggest that two-step TOR procedures are in fact superior to one-step TOR procedures with regard to preference ratings. Two-step TOR procedures have the potential to create a more gradual take-over process, by providing a time frame between warning and alarm, within which the driver can take-over control. Therefore, they are suited for uncritical take-over situations with enough time available to switch control (e.g. at the end of a phase
of highly automated driving on an autobahn). However, two-step TOR procedures might not be applicable in highly dynamic situations where driver reactions have to follow the TOR promptly. Future studies should examine the applicability of two-step TOR procedures in varying driving situations.

Concerning the second objective, our results indicate that a visual TOR consisting of text and pictograms was rated as more intuitive and attractive than a TOR consisting of text only. Moreover, a visual TOR consisting of text and pictograms was rated as more useful and appropriate in displaying the information in the auditory condition with a single tone. This effect could not be found in the auditory condition with speech. These findings suggest that adding pictograms to a text can be more intuitive and aesthetic but it does not increase the usefulness of a TOR containing speech any further. During periods of highly automated driving, the drivers’ visual attention is likely to be involved in secondary tasks or in monitoring the driving situation. His or her capacity to process additional visual information of the TOR might be very limited. Hence, dashboard displays should not contain excessive verbal information. Instead, verbal information could be conveyed via the auditory channel. Our findings show that an auditory TOR consisting of speech was rated as more intuitive and more useful than a single tone. Moreover, speech outranked pure tone within two-step and one-step procedure conditions. These results are in line with Forster et al. (2017), who found auditory TORs containing verbal information to be superior to non-verbal TORs for one-step TOR procedures. Future studies should further investigate the potential of non-visual car-to-driver interaction (Gellatly, Hansen, Highstrom, & Weiss, 2010). Pictograms, text, and beeps might no longer be the primary way of communicating information from vehicle to driver. Instead, the interaction of drivers and vehicles should be based on meaningful auditory information.

The current study examined subjective preferences of TOR interfaces. It demonstrated that a TOR design containing a two-step procedure and verbal auditory information (speech) is preferred by drivers. However, shortcomings of the study should also be noted here. Firstly, the study was conducted as an online study. Thus, there was no controlled testing environment (e.g. screen size), no real driving experience, and no assessment of driving performance. Future studies should apply the TOR interfaces in real driving scenarios. Moreover, the manipulation of TOR modalities was not exhaustive. Future studies could investigate the impact of non-verbal visual (pictogram only) and hybrid auditory (speech and tone) TORs on drivers’ preferences. Despite the shortcomings of the study, the current investigation addresses an important aspect in the development of highly automated vehicles. The success of automated vehicles partially depends on their human-machine interfaces (Casner et al., 2016). The open question on how to design HMI for take-over requests in highly automated driving is therefore one of the very important research questions in the developing research field. The present study added some information on the design of human-machine interfaces. It highlighted the value of two-step TOR procedures and verbal auditory information when issuing take-over requests. Future studies could examine their characteristics and how they influence driver behaviour.

REFERENCES


Evaluation of free public transport for older people in Sweden

Tania Dukic Willstrand\textsuperscript{1,2}, Sweden, tania.willstrand@vti.se, Per Henriksson\textsuperscript{1}, Sweden, Helena Svensson\textsuperscript{2,3}, Sweden, Lena Levin\textsuperscript{1,2} Sweden.

\textsuperscript{1}VTI – The Swedish National Road and Transport Research Institute, SE-581 95 Linköping, Sweden.
\textsuperscript{2}K2 – The Swedish Knowledge Centre for Public Transport, Scheelevägen 2, SE-223 81 Lund, Sweden.
\textsuperscript{3}Lund Technical University, Sweden.

ABSTRACT

Older citizens safe mobility is an issue as the number of older people is growing and expect to live longer than previous generations. To keep their independence and to allow them to take part in the society, transport accessibility is an issue to be solved. The present study developed a survey to evaluate a subsidised public transport card for older citizens in the western part of Sweden and how the physical health contributes to the use of public transport. A questionnaire was sent to 1500 older citizens in three municipalities to examine how this measure influenced their travel patterns and whether this is an efficient measure to increase their use of public transport. Results show a significant effect of the senior card which vary depending on the municipality and incomes. Some older citizens increased travelling with public transport (PT), they changed the time of the day for some activities and to some extend prioritized PT in another way than before they got the subsidised card. The senior card contributed to shift travel mode choice from private car to PT but also from cycling/walking to PT.

To reach a sustainable safe mobility for older citizens, a discussion is on-going to find and target measures to this broad group of individuals. Health and environment goals need to be considered to reach the desired results.

Keywords: older, public transport, free senior card, car driving.

9 INTRODUCTION

The number of years people expects to live is increasing (European Commission 2011). In addition, the number of years with good health is increasingly fast. This new paradigm of “Ageing Society” or “Long healthy life” is having a significant impact and strain on our society (WHO. 2002). In the future, requirements from the transport system will be crucial firstly, to allow people to keep their independence, and secondly to allow them to take part in the society and to keep their social network (Owsley 2002). Keeping older people mobile in later life is decisive to sustain their autonomy, which has a significant impact on a social and economic perspective. Older citizen will drive more years in older age (Koppel and Berecki-Gisolf 2015), but they will also face specific problems while driving at old age. Therefore, society needs to be proactive and increase knowledge on how to early attract future older people to the public transport since they sooner or later will need it and be dependent on it (Fiedler 2007). Barriers need to be identify.

Several municipalities in Sweden and countries in Europe offer free public transport trips to older citizens through some form of “Senior card” (Laverty and Millett 2015). Rules for obtaining such a card differ in different municipalities regarding age (65+ or 75+) and regarding the time of the day to use it. Monetary incentives to increase public transport use has been reported earlier to have several effects such as increase of daily motion and social interaction (Webb, Laverty et al. 2016). However, research is not unambiguous regarding those effects.
Our theoretical framework is based on the capability concept (Sen 2009), i.e. people ability to reach their goal and make things that the perceived as valuable. In aging, these opportunities are affected by health, genetics, personality, cognitive ability, family, friends, housing, etc. In terms of mobility aspects, human ability to travel is influenced by the transport system design, costs and their accessibility. In the present study, capability is used to study how the physical capability influence the willingness and the actual use of public transport. Physical ability is measured here by asking respondents how long they could walk outside without any help and how often they actually walked.

1.2. Objectives

The present study aims to evaluate the effect of subsidised public transport for older citizens in three municipalities in the western part of Sweden. The effects are studied based on how older use of public transport has changed due to the introduction of a senior card. A special focus is to examine the relationship between subsidised public transportation and physical capability.

2. METHOD

2.1. Participants

A random sample from the Swedish Population register (SPAR 2017) was done for 250 women and 250 men from each of the three municipalities. The only criterium was that citizen had to be older than 65 years old in 2017 to fulfil the senior card requirement. Two municipalities were chosen based on a broad range of public transport and one municipality with a limited range of public transport (Table 1).

<table>
<thead>
<tr>
<th>Municipality</th>
<th>Environment</th>
<th>Public transport density</th>
<th>Conditions for use of the Senior card</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Göteborg</td>
<td>Urban</td>
<td>High</td>
<td>Low traffic hours (8:30-15 &amp; 18-06), 65 years</td>
<td>250 men &amp; 250 women</td>
</tr>
<tr>
<td>Mölndal</td>
<td>Suburb</td>
<td>High</td>
<td>Low traffic hours (8:30-15 &amp; 18-06), 65 years</td>
<td>250 men &amp; 250 women</td>
</tr>
<tr>
<td>Svenljunga</td>
<td>Rural</td>
<td>Low</td>
<td>24h/everyday, 65 years + one-time fee of 15€</td>
<td>250 men &amp; 250 women</td>
</tr>
</tbody>
</table>

2.2. Survey

A survey was designed to evaluate the senior card among older citizen. The survey was composed by 43 questions in total. The questions covered participants’ background, health, travel habits, use of public transport and senior card, car driving, travel experience and everyday life satisfaction. The survey was sent by post on October 20, 2017. No reminder was sent. Completed questionnaires were scanned and the result was delivered to VTI for further analysis. No personal data was collected that could identify the respondents. Statistical analyses were done with SPSS® (version 22.0). A p-value of <0.05 was considered statistically significant. Pearson Chi-square tests were used for non-parametric data analysis.
10 RESULTS

The response rate was 43%, i.e. 648/1500 participants. 45 percent of men answered the survey respectively 42 percent of women. The mean age was 75 years [66-93]. About 1/3 lived alone and 63% lived in a relationship. Concerning their living conditions, 45% lived in a flat and 53% in their own house.

3.1. Senior card users

In total, 80% of the participants who received a senior card offer did accept it, 13% did not. Within the 80%, 64% do use the card all the time or very often. A significant difference in card usage was found between different municipalities where users are mostly found in urban and suburban areas ($\chi^2=264; p<0.01$), see Figure 1.

3.2. Changes in travel patterns

There is a general effect in terms of PT use increase after the introduction of the senior card, 61% of users reported an increase of PT use after receiving the senior card (Figure 2). However, the effect is different depending on the municipality where they lived (Göteborg 67%, Mölndal 56% and Svenljunga 45%; $\chi^2=19.1; p<0.01$) and depending on the incomes where the less incomes the more use of PT ($\chi^2=32.2; p<0.01$). About half of the users has adjusted the time slot to use PT to fit into the senior card traffic hours. Both municipality and incomes do have an effect where users living in an urban area and with lowest incomes adjust their time the most. About 1/3 users reported the card to be too limited in terms of geography and hours to satisfy their travel needs.
To examine whether the senior card contributes to a shift of transport mode, card users were asked how they travelled before and after the senior card introduction for their different activities. Overall, there is a consistency of the majority of transport mode before and after the senior card for most of the travels (Table 2). For travels done by PT before the senior card introduction, 97% of the travels in average are still done by means of PT after the introduction of the senior card. For travels done with a car before, a shift in favour to PT is observed for service, meet friends/family and associations activities. PT afterwards account for 24-35% of these trips. For travels done with a cycle or by walk before, a shift in favour to PT is observed for shopping, service and meet friends/family. The corresponding PT percentage is now 30-35%.

Table 2: Shift of transport mode after the introduction of the senior card.

<table>
<thead>
<tr>
<th>Travels done by (bus/car/cycle) before the senior card</th>
<th>Travels done by public transport after the introduction of the senior card</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Bus" /> 97%</td>
<td><img src="image" alt="Bus" /> 97%</td>
</tr>
<tr>
<td><img src="image" alt="Car" /> 24 - 35%</td>
<td><img src="image" alt="Bus" /> 97%</td>
</tr>
<tr>
<td><img src="image" alt="Cycle" /> 30–35 %</td>
<td><img src="image" alt="Bus" /> 97%</td>
</tr>
</tbody>
</table>
3.3. Physical capability and transport

Physical capability and transport choice was studied by examining the relationship between the introduction of the senior card (i.e. if respondents have applied for it or not) and how long the respondents could walk outside without any help (i.e. from 0-200m to more than 1km). Respondents who started to use the senior card were significantly more able to walk long distances outdoors, compared to those who did not use the senior card ($\chi^2 = 8.4; p<0.05$). Moreover, respondents who used the senior card were more satisfied with their possibilities to travel with PT ($\chi^2 = 264; p<0.01$). In contrast, respondents who did not accept the senior card are the one who are driving almost every day.

11 DISCUSSION

The present study shows a significant effect of the senior card onto travels patterns of older drivers. These effects are varying depending on the municipality of living, the economical situations as well as the household composition. Majority of the senior card users reported to have increase their PT use, half of them reported an adjustment of their hours to use PT to fit the card requirements. These changes primarily concern people in urban and suburban areas, with lower incomes and for services and social activities. Regarding the travel pattern of the senior card users, a shift of 24-35% (depending on activities) of travels was observed from card to PT and 30-35% from cycle/walk to PT. Although, a significant relationship was found between respondents physical capability and the use of the senior card where the better the physical capacity the more use of the senior card.

The senior card seems to have contribute to decrease to some extend the number of travels done by car in favour of PT use. Earlier research has also showed a relationship between the use of (subsidised) PT and benefits in terms of increase physical activity (Coronini-Cronberg, Millett et al. 2012, Webb, Laverty et al. 2016, Rouxel, Webb et al. 2017). The shift from cycling and walking to PT is quite common when PT service becomes fully subsidised, not only for older age groups. Studies have shown that although the measure gives a major travel increase, free public transport might also contribute negative environmental and health effects due to supply increase and shift from bicycle and walking to public transport (Fernley 2013, Nilsson, Stjernborg et al. 2017).

In conclusion, the measure to sponsor PT travels for older citizens seems to have a rather positive effect regarding their mobility and their level of physical activity. However, the effect seems to be limited to older people who are rather fit physically and live in areas where there is a rather good availability of PT. A holistic approach is needed to cover a broader spectrum of older citizens and to insure that their needs are covered by the available public transport.

12 ACKNOWLEDGMENTS

Thanks to the respondents who took their time to answer the actual survey. This research has been funded K2, the Swedish Centre for Research and Education on Public Transport in Sweden and by the region Västra Götaland.

13 REFERENCES


DISTRACTION AND INNATENTION
Parallel session
Self-regulation of Drivers’ Mobile Phone Use: The Influence of Driving Context

Michiel Christoph, SWOV, the Netherlands, michiel.christoph@swov.nl, Simone Wesseling, SWOV, the Netherlands, Nicole van Nes, SWOV, the Netherlands

ABSTRACT
Mobile phone use while driving is considered as a major concern for traffic safety. Various studies indicate negative effects of distracted driving and recent Naturalistic Driving studies report substantial increases in crash risk of mobile phone use while driving. The objective of this study was to investigate what mechanism related to self-regulation underlies drivers’ decision to engage in mobile phone activity while driving. This study focused on the effect of driving context. For this study naturalistic driving data collected in the UDRIVE project was analysed. Dutch drivers spent over 9% of all driving time engaging in mobile phone related tasks. Drivers used their mobile phone significantly less when a passenger was present. Also a significant overrepresentation of visual-manual (VM) tasks initiated during standstill was observed, for the other speed categories significantly less VM tasks than expected were initiated. In addition, significantly more time was spent engaged in VM tasks on urban roads than expected. On rural roads and highways significantly less time was spent on VM tasks than expected. The analysis clearly shows indications of drivers’ self-regulatory behaviour.

Keywords: Distraction, Mobile phone, Driving context, Self-regulation, Naturalistic Driving.

1 INTRODUCTION
In 2017, 89% of all Dutch inhabitants of twelve years and older own a smartphone (CBS, 2018). The rapid penetration and use of mobile phones in the last decade generated a wide interest in safety issues related to mobile phone use while driving (SWOV, 2017). Various (review) studies investigated the behavioural effects of the use of a mobile phone while driving (Basacik, Reed & Robbins, 2011; Collet, Guillot & Petit, 2010a; 2010b; Dingus, 2016; Stelling-Konczak & Hagenzieker, 2012). Based on the results of behavioural studies, we can conclude that using a mobile phone has a negative effect on driving behaviour indicating that mobile phone use while driving is a problem for traffic safety. In order to understand the magnitude of this problem better, this study investigated the prevalence of mobile phone use of Dutch car drivers. In addition, driving context when using the mobile phone was analysed in order to explore mechanisms of self-regulatory behaviour.

In the last decade several studies using different methodologies researched prevalence of mobile phone use while driving. Recent studies using Naturalistic Driving (ND) data show large differences in prevalence of mobile phone use between countries. A recent US study based on ND data reports (hand-held) mobile phone use of over six percent of all driving time (Dingus et al., 2016). Results of the recent European ND project UDRIVE show large differences between different European countries (Carsten et al., 2017) ranging from below one percent to above nine percent of all driving time. While these differences remain largely unexplained yet, it highlights the importance of obtaining national or regional data on prevalence of mobile phone use.

Mobile phone-related accidents have not increased in line with the use of the mobile phones suggesting that the potential risks of mobile phones use are regulated at many levels (Pöysti, 2005). Drivers self-regulatory behaviour of mobile phone use while driving can occur at different levels: strategic level (e.g. deciding not to use a mobile phone while driving), tactical level (e.g. the timing of engagement in the mobile phone task) or at the operational level (e.g. slowing down, often referred to as compensatory behaviour). This study aims to explore how driving context influences the
drivers’ decision to use the mobile phone on a tactical level. The focus will be on visual-manual (VM) phone tasks such as dialling, sending a text message or reading as they are commonly associated with increased risk (Klauer et al., 2006; Olson, Hanowski, Hickman & Bocanegra, 2009). Holding a mobile phone (including hand-held calling, texting) while driving is prohibited by law in the Netherlands, hands-free calling however is allowed.

2 METHOD

2.1 The UDRIVE database of Dutch car drivers
Naturalistic driving data of Dutch drivers collected in the UDRIVE project between 2015 and 2017 was used for analysis in this study. Thirty-three drivers participated in the Dutch car study; 3727 hours of data was collected and in total 230,842 kilometres were driven by the drivers. Data of 28 participants were included in this study. The average age of participants was 44.5 years (SD=12.9; range 26-70), of one participant age was unknown. An equal amount of women and man participated.

All participants were provided with a leased Renault Clio IV which they drove in during their participation. The cars were equipped with seven cameras and a data acquisition system able to log CAN-bus data, GPS-data, Mobileye data and video footage (more information on the UDRIVE data collection see Bärgman et al., 2017). In this study video images were analysed to observe mobile phone usage. Map matched GPS-data and CAN-bus data were used to determine driving speed, road type and speed limit.

2.2 Data sample and annotations
A random sample of trips was selected of the 28 participants out of the available data. Two inclusion criteria for trip inclusion were used; a trip duration was at least three minutes and 50% of the trip duration driving speed had to be above 5 km/h. A team of five video annotators manually inspected trips from start to end for episodes of mobile phone use. The following subtasks were annotated: hand-held conversation, hands-free conversation, hand-held interaction, hand-held reading, hands-free interaction, holding (without interaction), searching for the mobile phone and any other activities related to the mobile phone (other). The presence of passengers was annotated as well. When the quality of the inward camera was not sufficient the trip was excluded and a different random trip was selected. Consequently, 656 trips were annotated and 225 hours of video material, including 14159 driven kilometres of 28 participants. Per participant a minimum of 10 and a maximum of 36 trips were annotated.

2.3 Data analysis
Prevalence of mobile phone engagement, the effect of passenger presence and the influence of speed and driving context on the engagement in VM tasks were analysed in detail. VM tasks were defined as the subtasks hand-held interaction and hand-held reading. If there were less than three seconds between VM tasks, the VM task was annotated as one.

3 RESULTS

3.1 Prevalence of mobile phone engagement
Prevalence of the aforementioned categories of mobile phone engagement was determined per participant. Averages over participants were then calculated. In 32% (SD=25.6) of the trips a mobile phone was used. Differences amongst participants were large. One participant never used the mobile phone in the car while another participant used the mobile phone in all of the trips that were annotated. Drivers spend 9.3% of all driving time engaged in mobile phone related tasks. One driver
was even engaged in mobile phone related tasks during 74% of the driving time. Table 1 gives an overview of the percentage of time that was spend on a specific task related to mobile phone use. Drivers rarely called hand-held, while hands-free calling happened more often, 2.1% of all driving time. Drivers were engaged in a VM task (hand-held interaction and hand-held reading) in 1.7% (SD=2.5) of all driving time.

Table 1 – Average, standard deviation and maximum percentage of the time wherein drivers are engaged in mobile phone related tasks

<table>
<thead>
<tr>
<th>Task</th>
<th>Average percentage</th>
<th>Standard deviation</th>
<th>Maximum percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total phone engagement</td>
<td>9.3</td>
<td>16.1</td>
<td>73.7</td>
</tr>
<tr>
<td>Hand-held conversation</td>
<td>0</td>
<td>0</td>
<td>0.02</td>
</tr>
<tr>
<td>Hands-free conversation</td>
<td>2.1</td>
<td>3.1</td>
<td>12.2</td>
</tr>
<tr>
<td>Hand-held interaction</td>
<td>1.5</td>
<td>2.2</td>
<td>10.4</td>
</tr>
<tr>
<td>Hand-held reading</td>
<td>0.2</td>
<td>0.4</td>
<td>1.8</td>
</tr>
<tr>
<td>Hands-free interaction</td>
<td>0.6</td>
<td>1.5</td>
<td>7.1</td>
</tr>
<tr>
<td>Holding (without interaction)</td>
<td>4.5</td>
<td>12.3</td>
<td>56.8</td>
</tr>
<tr>
<td>Searching for the mobile phone</td>
<td>0.2</td>
<td>0.2</td>
<td>0.7</td>
</tr>
<tr>
<td>Other</td>
<td>0.02</td>
<td>0</td>
<td>0.1</td>
</tr>
</tbody>
</table>

The effect of passenger presence was determined with a paired-sample t-test. The percentage of trips with passengers present was compared with trips wherein the mobile phone was used and passengers were present. In 26% of all the trips (SD=18.3) a passenger was present during the major part of the trip. For trips where a mobile phone was used this was 16% (SD=23.8), significantly less than the overall percentage ($t(26)=2.973$, $p=0.006$, $d=0.57$). This indicates that drivers used their mobile phone less often when there was a passenger present. When only looking at visual-manual interactions in 6.9% (SD=19.6) of the trips where a mobile phone was used a passenger was present, also significantly less than the overall percentage of 26% ($t(26)=7.767$, $p<0.001$, $d=-1.493$).

3.2 Visual-manual tasks, speed and driving context

1058 visual-manual tasks were identified, drivers were engaged in a VM task 3.9 hours of the driving time. The speed with which drivers initiated a VM task was analysed. To do so, the percentage of VM tasks initiated at a certain speed was compared to the overall percentage of time driven at a certain speed in speed bins of 10 km/h (conform Tivesten and Dozza, 2015). The results are shown in Figure 1. Clearly, most of the tasks were initiated when standing still.

Figure 1 - The percentage of driving time in speed bins of 10 km/h for all annotated trips (225 hours) and the percentage of visual-manual tasks initiated at a certain speed
To gain understanding of the relation between speed driven and VM task engagement, the number of observed VM tasks initiated in a certain speed bin was compared to the expected number of VM tasks based on the distribution of driving time in a certain speed bin. Figure 2 shows the difference between the observed and expected values. An overrepresentation of VM tasks initiated during standstill was observed, for the other speed categories an underrepresentation was observed. A chi-square goodness of fit test was used to compare the occurrence of VM tasks with the hypothesized occurrence. Significant deviations from the hypothesized values were found for all categories (Standstill: $\chi^2(1)=358.1$, $p<0.001$; 0-50 km/h: $\chi^2(1)=11.7$, $p<0.001$; 50-100 km/h: $\chi^2(1)=22.9$, $p<0.001$; >100 km/h: $\chi^2(1)=46.2$, $p<0.001$).

The data was further analysed to explore the relation between road type and engagement in VM tasks, focussing on urban areas (30-50 km/h speed limit), rural areas (60-90 km/h speed limit) and highway (>100 km/h speed limit). For 74% of the visual-manual tasks (787 tasks) speed limit information was available, corresponding to 2.1 hours of the VM tasks (53% of the total duration of VM tasks). For these tasks the percentage of time driven on a road with a certain speed limit was calculated. Most of the time engaged in a VM task was spend on highways (50%), 17% of the time engaged was spend on rural roads and 33% of the time engaged was spend on urban roads. As drivers do not spend an equal amount of time on the different road types, we again compared the observed and expected time spend on VM tasks based on the distribution of driving time on the different road types. A chi-square goodness of fit test was used to compare the total duration of VM tasks per road category with the expected total duration of VM tasks per road category. Significant deviations from the hypothesized values were found for all categories (urban: $\chi^2(1)=155.6$, $p<0.001$; rural: $\chi^2(1)=20.9$, $p<0.001$; highway: $\chi^2(1)=55.0$, $p<0.001$). The results presented in 3 show that given the distribution of time driven on a certain road type, more time was spend engaged in VM tasks in urban roads than would be expected, on rural roads and highways less time was spend than would be expected.
DISCUSSION AND CONCLUSION

On average, Dutch drivers spend 9% of all driving using their mobile phones. Prevalence in the Netherlands is high, compared to the prevalence of mobile phone use in other European countries determined using the same methodology (Carsten et al., 2017): 3.5% in France, 1% in Germany, 9.8% in Poland and 2.9% in the UK. Further analysis is needed to explain these large differences. Methodological aspects such as participant sample could partly explain the differences, but also cultural differences, differences in legislation and enforcement could be related to the observed results. Better understanding of the observed differences between European countries could facilitate the development of countermeasures and support decision making.

Drivers used their mobile phone significantly less when a passenger was present. Also a significant overrepresentation of VM tasks initiated during standstill was observed, for the other speed categories significantly less VM tasks than expected were initiated. In addition significantly more time was spend on VM tasks on urban roads than expected. On rural roads and highways significantly less time than expected was spend on VM tasks.

Although the results presented in this paper are limited and the participant sample cannot be considered as representative for the Netherlands, there are clearly some indications of drivers’ self-regulatory behaviour. This paper describes the intermittent results of a currently ongoing analysis. More factors of driving context such as driving manoeuvre, road curvature, weather/light conditions and lead vehicle presence will be included in the analysis. In addition, a 15-second baseline period preceding the engagement of a VM task will be compared to driving during VM task engagement on several factors.

The majority of drivers are likely to be aware of the dangers of mobile phone use behind the wheel, partly because of public campaigns. It would be a reasonable assumption that one of the drivers’ goals is to drive safely and avoid crashes. If self-regulatory behaviour related to mobile phone use of drivers would be perfect, this goal would be reached and no mobile phone use related accidents would happen. The literature describes different patterns of self-regulation failure such as underregulation (deficient standards, inadequate monitoring, or inadequate strength) and misregulation (false assumptions or misdirected efforts) (Baumeister & Heatherton, 2009). Self-regulatory behaviour related to mobile phone use of drivers placed against this theoretical background helps understanding drivers’ decisions to engage in mobile phone activity while driving but most importantly it increases understanding of self-regulation failure that leads to unsafe driving.

Figure 3 - The difference between observed and expected duration of VM tasks in seconds in urban, rural and highway driving context (N=787)

4 DISCUSSION AND CONCLUSION

On average, Dutch drivers spend 9% of all driving using their mobile phones. Prevalence in the Netherlands is high, compared to the prevalence of mobile phone use in other European countries determined using the same methodology (Carsten et al., 2017): 3.5% in France, 1% in Germany, 9.8% in Poland and 2.9% in the UK. Further analysis is needed to explain these large differences. Methodological aspects such as participant sample could partly explain the differences, but also cultural differences, differences in legislation and enforcement could be related to the observed results. Better understanding of the observed differences between European countries could facilitate the development of countermeasures and support decision making.

Drivers used their mobile phone significantly less when a passenger was present. Also a significant overrepresentation of VM tasks initiated during standstill was observed, for the other speed categories significantly less VM tasks than expected were initiated. In addition significantly more time was spend on VM tasks on urban roads than expected. On rural roads and highways significantly less time than expected was spend on VM tasks.

Although the results presented in this paper are limited and the participant sample cannot be considered as representative for the Netherlands, there are clearly some indications of drivers’ self-regulatory behaviour. This paper describes the intermittent results of a currently ongoing analysis. More factors of driving context such as driving manoeuvre, road curvature, weather/light conditions and lead vehicle presence will be included in the analysis. In addition, a 15-second baseline period preceding the engagement of a VM task will be compared to driving during VM task engagement on several factors.

The majority of drivers are likely to be aware of the dangers of mobile phone use behind the wheel, partly because of public campaigns. It would be a reasonable assumption that one of the drivers’ goals is to drive safely and avoid crashes. If self-regulatory behaviour related to mobile phone use of drivers would be perfect, this goal would be reached and no mobile phone use related accidents would happen. The literature describes different patterns of self-regulation failure such as underregulation (deficient standards, inadequate monitoring, or inadequate strength) and misregulation (false assumptions or misdirected efforts) (Baumeister & Heatherton, 2009). Self-regulatory behaviour related to mobile phone use of drivers placed against this theoretical background helps understanding drivers’ decisions to engage in mobile phone activity while driving but most importantly it increases understanding of self-regulation failure that leads to unsafe driving.
REFERENCES

https://doi.org/10.26323/UDRIVE_D41.1


https://doi.org/10.26323/UDRIVE_D43.1


What Are Drivers Doing When They Aren’t on the Cell Phone?

Carol Flannagan, University of Michigan, USA, cacf@umich.edu, Jonas Bärgman, Chalmers University of Technology, Sweden, András Bálint, Chalmers University of Technology, Sweden

Keywords: Driver distraction, crash risk, cell phone

ABSTRACT

Cell-phone bans have been motivated by previous research estimating large odds ratios (i.e., 3-4) for using cell-phones while driving. However, large crash reductions have not been realized. One reason may be that drivers may replace cell-phone use with other risky activities and that ORs have often compared cell-phone use to ideal driving rather than a realistic reference. Using SHRP2 data, we developed two cell-phone propensity models, one with age and one without, to develop weights for events without cell phone use. Using these weights, we estimated the probability of engagement in a variety of tasks in place of cell-phone use. We also estimated weighted ORs for cell-phone use (all uses) and cell-phone talking only. Weighted ORs are lower than unweighted ORs and much lower than ORs compared to ideal driving. This is consistent with the idea that in practice, even if cell-phone bans are effective at reducing cell-phone use, they may not greatly reduce risk because drivers may replace cell-phone use with other distracting activities in the same situations in which they normally use cell phones while driving. We also discuss the influence of young drivers on our results. Younger drivers in the dataset are more likely to use cell phones and thus are influential in the propensity model results.

ACKNOWLEDGEMENTS

We would like to thank the Transportation Research Board and the National Academies for making the SHRP2 data available for analysis. The data used has the unique DOI: 10.15787/VTT1/EYG6RB. The findings and conclusions of this paper are those of the author(s) and do not necessarily represent the views of VTTI, the Transportation Research Board, or the National Academies.

INTRODUCTION

Previous research using the Second Strategic Highway Research Program (SHRP2) Naturalistic Driving Study (NDS) reported that the estimated odds ratio (OR) of crashing associated with cell-phone use while driving was 3.6 (CI: 2.9-4.5), when compared to ideal driving, in which drivers were not fatigued, impaired or engaging in any secondary tasks (Dingus et al., 2016). Related work using crash data (e.g., McEvoy et al., 2004) has led to cell-phone bans being pass in various states and countries.

In spite of the evidence for risks associated with cell-phone use, the effect of bans on crashes has not been realized. In the U.S. an analysis of states that banned cell-phone use when driving showed that the bans reduced cell-phone use but not insurance claims (IIHS, 2014). In Europe, the UDrive study found large differences in mobile-phone use in driving by country (Carsten et al., 2017). German drivers used mobile...
phones 0.06% of the time as compared to UK (2.87%), French (3.48%), and Polish (9.49%) drivers. All of these countries have mobile-phone bans while driving though very different rates were observed in the data.

One hypothesis for the failure to realize large reductions in crash counts where mobile-phone bans are in place is that drivers who are not using their cell phones may engage in difference secondary tasks and be distracted by other means. Since NDS-based OR estimates for cell-phone use are often calculated using ideal driving as a reference activity, it may be that these ORs overestimate the potential benefits that can be realized in actual driving when cell-phone use is reduced. That is, drivers are unlikely to replace cell-phone use with ideal driving, since they are engaged in secondary tasks frequently (more than half the time in SHRP2; Dingus et al., 2016). The question is: What is an appropriate reference activity that better reflects the likely activities with which drivers will replace cell-phone use? In addition, what is the OR for cell-phone use when compared to the likely replacement activity?

In this paper, we approach this problem by developing a new reference in which non-cell-phone events are weighted by the probability that the driver would have used a cell phone in that event, given the context present. Here, context is defined as environmental characteristics (e.g., traffic density, time of day) and driver characteristics (age, gender). This approach is similar to propensity weighting (Freedman & Berk, 2008; Månsson et al., 2007) in that we create a cell-phone-use propensity model that is applied to each event. However, propensity weighting is typically used to correct for overrepresentation of certain cases (based on their covariates) in the treatment group relative to random assignment. Here, we have a random sample of all driving but want it to be a random sample of the driving conditions that are present when drivers are on the cell phone. Thus, we weight events by their cell-phone-use propensity. In this way, cell-phone crash odds are compared to reference driver behaviour and crash risk should better reflect the expected behaviour and crash risk when cell-phone use is discouraged (e.g., by legislative or technological) means.

The two objectives of this study are: 1) identify what drivers are doing when they are not on the cell phone, and 2) estimate the OR for cell-phone use when compared to the set of activities that would be expected to replace cell-phone use in the case of a ban.

**METHOD**

For this analysis, we used data from the SHRP2 study. SHRP2 is a naturalistic driving study (NDS) that was conducted in 6 locations in the U.S. from 2012-2013. Approximately 3,400 vehicles were monitored with Data Acquisition Systems (DASs), which collected video and kinematic data continuously while the vehicles were driven. This resulted in a total of approximately 5.5 million trips with consented drivers (i.e., study participants, whose data are available for analysis) and approximately 56 million kilometres of driving (Hankey et al., 2016).

As part of the original study, a set of safety-critical and non-safety-critical (baseline) events were extracted from the larger dataset. Safety critical events (SCEs) in the dataset include crashes (as well as near-crashes, which we did not use). In addition, a set of clips, known as “balanced baselines,” were selected at random from all driving over 5 mph. The number of clips selected per driver was proportional to his/her total driving time in the study.

Crashes were defined as:
Any contact that the subject vehicle has with an object, either moving or fixed, at any speed in which kinetic energy is measurably transferred or dissipated is considered a crash.”
(Hankey et al., 2016)

In addition, crashes were further categorized into Level I, II, III, and IV, in order of decreasing severity, based on video review. Level I is a severe crash, Level II is a moderate crash, Level III is a minor crash, and Level IV is defined as a low-risk tire strike. Detailed definitions and extraction methods can be found in Hankey et al., (2016).

For this analysis, we only used those crashes where the first event was coded as a Level I, II or III crash. Each baseline or crash clip was 20-30 seconds long but was only video-coded for 6 seconds. Video coding was done for the 5 seconds prior and 1 second after the identified precipitating event for crashes and for a random 6 seconds for baselines. This resulted in a dataset with 19,991 balanced-baseline cases and 830 crashes.

We developed our cell-phone propensity model using logistic regression to predict the probability of using the cell phone as a function of available covariates, which included:

- Relation to junction—spatial relationship of vehicle to a junction, if any
- Intersection influence—judgment of whether an intersection influenced driving behaviour such as braking
- Traffic flow—roadway design, including number of lanes and presence of dividers
- Traffic density—traffic density for the specific event, coded from free-flow to flow restrictions (stop-and-go)
- Front seat passengers—codes presence or absence of a front seat passenger
- Lighting—light level coded from forward video (e.g., daylight, dusk/dawn, dark but lighted, dark)
- Weather—weather conditions coded from forward video
- Surface condition—road surface condition coded from forward video (e.g., wet, dry, icy)
- Through travel lanes—number of lanes in driver’s direction of travel (does not include turn lanes)
- Day of week
- Locality—Coded surroundings e.g., residential, business, rural, etc.
- Mean travel speed during full event
- Driver age group—first group includes ages 16-19 (4 years) with 5-year groupings thereafter

All cell-phone interaction types, both with hands-free and hand-held phones, were included in analysis, though talking on a hand-held phone makes up the majority of cases (1656, or 92%).

Only baseline cases were used for the propensity model since if cell-phone use itself increases risk, we would expect crash cases to include cell-phone use with higher probability across all conditions. The model was then used to compute predicted probability of being on the cell phone for all cases in which the driver was not engaged in a cell-phone task. The 1803 events (including both crashes and baseline events) where cell phones were in use were given a weight of 1 and the 19018 non-cell-phone cases were weighted according to the predicted propensity of using the cell phone in that case.
To estimate the probability of other tasks (in lieu of cell phone), we used balanced baseline events where cell-phone use was absent. The full list of secondary tasks coded is available on the InSight\(^1\) website, but the most common tasks include talking to passengers, external distractions, talking/singing, etc. The weighted proportion of each alternative task (as well as no secondary tasks) occurring in that sample was used to estimate the probability that each alternative task would replace cell-phone use.

To estimate the OR for cell-phone use compared to the likely set of baseline activities, we also computed the cell-phone OR using the propensity-score weights for the reference group. The ORs were estimated using logistic regression with no covariates in the model. This is equivalent to the 2X2 table in Figure 1 below.

<table>
<thead>
<tr>
<th></th>
<th>Cell Phone Used</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Crash</td>
<td>a</td>
</tr>
<tr>
<td>Baseline</td>
<td>c</td>
</tr>
</tbody>
</table>

*Shaded cells are weighted in analysis

**Figure 1 Illustration of weighted odds ratio calculation**

The standard OR equation based on the cells in Figure 1 is given in Equation 1. The values a, b, c and d are cell counts. This is the calculation used in Dingus et al., 2016 where “no cell phone” (i.e., cells b and d) included only cases of ideal driving with no secondary tasks, fatigue or impairment.

\[
OR = \frac{a/c}{b/d}
\] (1)

Equation 2 shows how propensity weights are used in the same calculation. Here, we count all events where cell-phone use is absent (but there may be other secondary tasks) in cells b and d, and then we weight them as described in Equation 2.

\[
OR = \frac{\sum w_a/\sum w_{ac}}{\sum w_b/\sum w_{bd}} = \frac{a/c}{b/d}
\] (2)

where \(w_i\) is the sum of weights for events counted in cell \(i\). Note that for cells a and c, the sum of weights is equal to the count of events, as shown on the rightmost side of Equation 2.

**RESULTS**

The dataset contained 3,539 unique drivers. The age distribution is shown in Figure 2 compared to the age distribution for licensed drivers in the U.S. Young drivers are substantially overrepresented and older drivers are also overrepresented but not to the same degree. Of the 19,991 baseline events, 1,656 (8%) included some type of cell-phone interaction by the driver. Of the 830 crashes, 147 (18%) included cell-phone interaction.

---

\(^1\) https://insight.shrp2nds.us/login/auth#/builder
We developed two cell-phone propensity models, one with and one without driver age group. Figure 3 shows the distributions of predicted probability of cell-phone use for baseline events in which the cell-phone was (1) and wasn’t (0) used for the two models. For the model without age (left), the mean predicted probability of cell-phone use was 0.0826 for events where the cell phone was not used and 0.0849 for events where the cell phone was used. The corresponding values for the model with age were 0.0796 (no cell) and 0.126 (cell used).

Using the two propensity weights, we first looked at the set of activities that drivers were engaged in when they were not on the cell phone, weighted by propensity. Using the unweighted baseline data, drivers were not engaged in any secondary task (ideal driving) 46.7% of the time. The corresponding propensity-weighted estimates are 47% (no age) and 43.2% (with age). Of the remaining events, Figure 4 shows the
unweighted and propensity-weighted distributions of task involvement for the model without age (gray) and with age (blue).

Figure 4 Task involvement proportions using unweighted and two propensity-weighted models

Odds ratios and confidence intervals were computed using logistic regression with no covariates. The results for the unweighted comparison to model driving, the unweighted comparison to task absent, and the two weighted estimates are shown in Table 1. Unweighted and weighted (with age) odds ratio estimates are also included for the cell-phone hand-held talking only task, for comparison.

Table 1 Simple odds ratio (OR) estimates for crash risk associated with all cell-phone use using different reference groups and cell-phone talk only using unweighted and weighted (with age)

<table>
<thead>
<tr>
<th>Reference Group</th>
<th>OR Estimate</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>All cell; Unweighted model driving</td>
<td>3.56</td>
<td>2.87-4.40</td>
</tr>
<tr>
<td>All cell; Unweighted task absent</td>
<td>2.38</td>
<td>1.97-2.86</td>
</tr>
<tr>
<td>All cell; Propensity weighted (no age)</td>
<td>2.45</td>
<td>1.80-3.39</td>
</tr>
<tr>
<td>All cell; Propensity weighted (with age)</td>
<td>1.98</td>
<td>1.48-2.70</td>
</tr>
<tr>
<td>Cell-phone HH talking only (unweighted)</td>
<td>1.29</td>
<td>0.89-1.81</td>
</tr>
<tr>
<td>Cell-phone HH talking only (weighted with age)</td>
<td>1.04</td>
<td>0.68-1.56</td>
</tr>
</tbody>
</table>

DISCUSSION

A great deal of research has looked at the increase in crash risk associated with cell phone use while driving. This has led to legislation in many states and countries, but large hoped-for reductions in crashes (especially fatal crashes) have not been realized so far. One possible reason for this is that the estimates of crash odds ratios for cell-phone use are based on comparisons to idealized driving that is not achieved even
when drivers are motivated not to use a cell phone while driving. In other words, drivers may replace cell-phone use with other secondary tasks that have similar risk profiles and thus not achieve the expected safety gains.

In this and previous work (e.g., Klauer et al., 2010), ORs using no secondary tasks or other forms of ideal driving are generally higher than those using “task absent” reference groups. We found that in the unweighted analysis, the confidence intervals of the two reference comparisons did not overlap. This is not surprising given that model driving is generally “safe” driving (i.e., attentive, sober).

The propensity-weighted models produced some interesting patterns. In particular, without age, we could not predict the situations under which drivers choose to use the cell phone. This conflicts with other published results, including one by the first author (e.g., Flannagan, Bao & Klinich, 2012), that showed evidence of self-limiting of cell-phone use based on context. If drivers were strongly context-sensitive in their cell-phone use, the context-only propensity model should have done a better job of differentiating events with and without cell-phone use. Instead, the best predictor of cell-phone use was age, suggesting that the biggest contributor is the driver, not the situation.

In keeping with the poor differentiation performance of the no-age-group propensity model, the results of task involvement and ORs using weights from that model look very similar to unweighted results. However, when age is included, some tasks become more likely to occur (in cell-phone-likely situations) including talking/singing, dancing, and adjusting the radio. In addition, the point estimate of crash OR for cell-phone use is lowest using this model. Also, the estimated OR associated with hand-held cell-phone talking only, went down from 1.29 to 1.04 (both not significantly different from 1).

If age is the best predictor of cell-phone use and the largest contributor to cell-phone related risk, it may be necessary to think about the problem differently. Young drivers have long been shown to be riskier than other drivers (Massie, Campbell & Williams, 1995), but cell phones have only been ubiquitous in driving for a decade. Moreover, using SHRP2 data, Guo et al. (2016) show that cell-phone associated crash risk was different for different age groups. Older drivers exhibited the most elevated crash OR for cell-phone use but rarely use the cell phone. Drivers from 16-29 also had higher ORs than middle-aged drivers but use the phone at higher rates. In the SHRP2 study, young drivers are overrepresented in the sample (relative to the population), are overrepresented in crashes (relative to other age groups), and are overrepresented in cell-phone use (relative to other age groups). Even in other studies, such as those using case-crossover designs and crash data, drivers under 29 make up a large portion of the sample (e.g., 48% in McEvoy, et al., 2005). Thus, some portion of the estimated crash risk may be “borrowed” from risk specific to younger drivers.

It is important to note that although the risks associated with cell-phone use may have been overestimated in many studies (and in the popular press), we do observe a significant estimated crash OR of 1.98 for all cell-phone use in the age-included propensity-weighted model. Given the cell-phone talking-only estimate of 1.04, the all-cell-phone-use risk must be associated with non-talking (or hands-free) cell-phone uses. Although the overrepresentation of young people in SHRP2 may still inflate the OR somewhat, we do not claim that there is no risk or cause for concern. Instead, we have tried to develop a more appropriate estimate of the potential effect of removing cell-phones from the driving population by comparing their use to likely alternative behaviours.
This work could be extended in a number of ways. First, while we did not observe evidence of context-sensitive cell-phone use, it may be that young drivers have not developed this sensitivity while older drivers have. A follow-on study could develop age-group-specific propensity models, which we did not. Another improvement would be to reweight the drivers in SHRP2 to be better representative of the driving population. Finally, repeating this analysis on other NDS datasets such as UDrive would lend further insight into the factors that influence drivers’ decision to use the cell phone when driving and improve estimates of crash ORs from cell phone use (and other distractions).

Although the SHRP2 study is large, the number of crashes, especially relatively serious crashes, is still very small. In addition, the study population overrepresents younger and older drivers (relative to the U.S. driver population in general) and relied on self-motivated volunteers. Thus, conclusions from SHRP2 may not fully generalize to the U.S. driver population. Moreover, since the study was conducted in the U.S., results may not fully generalize to the driver population in other countries.
REFERENCES


Effects of secondary Tasks and Display Position on Glance Behavior during partially automated Driving.

A. Hensch, Chemnitz University of Technology (TUC), Germany, ann-christin.hensch@psychologie.tu-chemnitz.de, N. Rauh, TUC, Germany, C. Schmidt, TUC, Germany, S. Hergeth, BMW Group, Germany, F. Naujoks, BMW Group, Germany, J. F. Krems, TUC, Germany, A. Keinath, BMW Group, Germany.

ABSTRACT

The driving task is becoming increasingly automated, thus changing the driver’s role. Moreover, in-vehicle information systems using different display positions and information processing channels might encourage secondary task engagement. During manual driving scenarios, varying secondary tasks and display positions could influence driver’s glance behavior. However, their impact on the driver’s capability to monitor the partially automated driving systems has not yet been determined. The current study assessed both the effects of different secondary tasks (Surrogate Reference Task (SuRT) vs. text reading) and display positions (head-up display (HUD) vs. center console) on driver’s glance behavior during partially automated driving. Participants engaged in several secondary tasks that were presented on different display positions while monitoring the partially automated system during a simulated car following task. Different automation system failures regarding the lateral and longitudinal control occurred while driving. A head-mounted eye-tracker recorded the participants’ glance behavior. Repeated measures ANOVAs revealed that the HUD yielded considerably longer eyes-on display time (total and mean glance durations) than the center console. Moreover, the text reading task resulted in longer total and mean glance durations than the SuRT. Similar to manual driving scenarios, the results showed a consistent effect of display position and secondary task on the driver’s glance behavior. Despite the longer eyes-on display time for the HUD, its proximity to the driving environment might enable a faster identification of and reaction to critical situations (e.g., due to system failures).

Keywords: partially automated driving, secondary task, eye tracking, head-up display, head-down display.

1 THEORETICAL BACKGROUND

In recent years, the driving task has become increasingly assisted (Flemisch, Kelsch, Löper, Schieben, & Schindler, 2008; Papadimitratos, de la Fortelle, Evenssen, Brignolo, & Cosenza, 2009), which has resulted in different levels of vehicle automation. Based on the Society of Automotive Engineers’ (SAE) taxonomy (SAE International, 2014), partial automation (i.e., Level 2 automation) takes over both the lateral and longitudinal control (e.g., by combining adaptive cruise control and automated steering). With this level of automation, the driver’s role changes from an active operator performing all aspects of the dynamic driving task to more of a passive system monitor (Merat, Jamson, Lai, & Carsten, 2012). Allocating some driving tasks to the vehicle’s automation (e.g., lateral and longitudinal control) has been hypothesized (Wierwille, 1993) and shown to reduce the driver’s workload (Ma & Kaber, 2005; Stanton, Young, & McCaulder, 1997; Young & Stanton, 2007). However, the extra cognitive resources now available to the driver due to partial automation use (Ma & Kaber, 2005), combined with a greater number of in-vehicle information systems (IVIS; Papadimitratos et al., 2009), might encourage...
Effects of secondary tasks and display position on glance behavior during partially automated driving.

the driver to engage in secondary tasks (Rudin-Brown, Parker, & Malisia, 2003). During fully manual driving, secondary task engagement has been shown to influence the driver’s glance behavior (NHTSA, 2012). For example, glance durations away from the road due to secondary task engagement increased with secondary task complexity (Victor, Harbluk, & Engström, 2005). Furthermore, IVIS are typically displayed in different positions (Knoll, 2015). During fully manual driving, different display positions have been found to influence the driver’s glance behavior (Hada, 1994). Previous research has noted significantly longer total and mean glance durations to a head-up display (HUD) than for different head-down displays (e.g., the center console; Ecker, 2013). However, the proximity between the driving environment and the HUD display position could reduce the effort required to shift attention between different areas of interest (AOI). Consequently, information access might be enhanced (Wickens, Goh, Helleberg, Horrey, & Talleur, 2003) despite the longer glance durations towards the HUD (Ecker, 2013). Thus, driving safety could be improved as the number of missed critical events decreases (Wierwille & Tijerina, 1996). Therefore, both the glance duration to a specific AOI and the AOI’s location in relation to the driving environment seem relevant to driving safety. Thus, automation requires even more attention regarding the human machine interface design (Lee, 2008). Nevertheless, prior research has not yet determined the impact different secondary tasks and display positions have on the driver’s capability to monitor the partially automated driving systems.

2 OBJECTIVES

The objectives of the current study were to assess the effects of different secondary tasks and various display positions on the driver’s glance behavior (i.e., eyes-on display time) during partially automated driving in a simulated environment. A methodological approach based on a previous study (Rauh et al., submitted for publication) was used to examine the effects of secondary task engagement (Surrogate Reference Task (SuRT; Mattes, & Hallén, 2009) vs. text reading) and display position (HUD vs. center console).

3 METHOD

2.1 Participants

A total of $N = 58$ participants were included in the study. Due to either system errors or simulator sickness, the data for only $N = 50$ participants (23 females, $M_{\text{age}} = 37.90$) could be analyzed. Participants were recruited via an announcement on the webpage of Chemnitz University of Technology. After completing an initial screening questionnaire, participants were selected based on their gender and age group. All participants received payment for their participation.

2.2 Material and procedure

Data was collected in a fixed-based driving simulator. A head-mounted Tobii Pro Glasses eye-tracker (Tobii AB, 2016) recorded participants’ glance behavior and the Tobii Pro Lab (Tobii AB, 2016) software was used to analyze the data. The simulated driving scenario was based on one used in a previous study (Rauh et al., submitted for publication). An 11 km long car following task on a highway was simulated while participants drove in a partially automated mode (i.e., the ego car took over the lateral and longitudinal control). The ego
Effects of secondary tasks and display position on glance behavior during partially automated driving.

car’s velocity (80 km/h) and following distance to the car ahead (of 70 m) was held constant (provided no automation failure occurred). No other vehicle traffic was included. Two types of system failures regarding either the a) lateral or b) longitudinal control occurred four times per drive (Figure 1; see also Lorenz & Hergeth, 2015). The types of system failure as well as the time onset of the failures were randomized. Only the first and third system failure occurrence (always including both lateral and longitudinal system failure) were analyzed to ensure that the occurring failure was not predictable by the participants. Participants were instructed to detect and respond to the system failures as quickly as possible (e.g., manually breaking when the ego car fails to react to the lead vehicle slowing down). After the failure occurred, the driver had to manually reactivate the system by pushing a button on the steering wheel.

Figure 1 – Types and order of system failures during each test drive. SF = system failures (adapted from Rauh et al., submitted for publication).

Additionally, participants performed several secondary tasks (i.e., SuRT, text reading, video task, manual radio-tuning task) on various display positions (i.e., HUD, center console, display behind the steering wheel, smartphone, separate display below the center one for the manual radio-tuning task). Only the results regarding two secondary tasks (SuRT and text reading) as well as two display positions (HUD and center console) will be reported in this paper (complete results will be presented elsewhere). The SuRT and text reading secondary tasks were chosen due to their opposite ecological validities (i.e., SuRT considered more artificial than the common task of text reading). During the SuRT, participants had to identify a target stimulus among several distractors (Petzoldt, Brüggemann, & Krems, 2014). The text reading task required participants to scroll down continuously to read the entire text. A control question followed each text to ensure that participants read the full text (no rewards were given for correctly answering questions). Participants practiced each secondary task twice prior to the respective test drive. The secondary tasks were conducted on two randomly assigned display positions, including a) the HUD and b) the center console, which differed by their distance to the road environment. Participants were previously instructed to prioritize the system monitoring task, which included scanning the surrounding driving environment. The glance behavior began being recorded (i.e., total and mean glance durations of eyes-on display time) one kilometer prior to the system failure, which served as an indicator of participants’ monitoring behavior.
4 RESULTS

Data were analyzed using repeated measures ANOVAs with type of secondary task and display position as independent variables, as well as total and mean glance duration on the respective display position as dependent variables. Effect sizes were interpreted based on Cohen’s recommendations (Cohen, 1988). A large effect ($F(1,37) = 14.87$, $p < .05$, $\eta^2_p = .29$) revealed longer eyes-on display time in terms of total glance duration for the text reading task ($M_{total} = 36.80s$, $SD_{total} = 4.77s$) than the SuRT ($M_{total} = 33.87s$, $SD_{total} = 6.77s$). Additionally, a medium effect ($F(1,36) = 2.31$, $p = .14$, $\eta^2_p = .06$) showed the mean glance duration was also longer for the text reading task ($M_{mean} = 7.24s$, $SD_{mean} = 9.69s$) compared to the SuRT ($M_{mean} = 4.76s$, $SD_{mean} = 5.42s$). Regarding display position, a large effect ($F(1,37) = 30.75$, $p < .05$, $\eta^2_p = .45$) indicated a considerably longer eyes-on display time in terms of total glance durations for the HUD ($M_{total} = 39.23s$, $SD_{total} = 4.25s$) in relation to the center console ($M_{total} = 32.18s$, $SD_{total} = 3.68s$). Further, a large effect ($F(1,36) = 20.80$, $p < .05$, $\eta^2_p = .37$) revealed that the mean glance durations to the HUD were longer ($M_{mean} = 9.42s$, $SD_{mean} = 6.88s$) compared to the center console ($M_{mean} = 2.37s$, $SD_{mean} = 0.73s$).

5 DISCUSSION

The current study investigated the impact of different secondary tasks and display positions on two analyzed glance parameters using a partially automated driving simulator. Text reading resulted in longer eyes-on display time compared to the SuRT. Furthermore, results indicated an effect of display position on driver’s glance behavior regarding the eyes-on display time. A large effect showed a difference between the two display positions regarding the eyes-on display time, which was based on the two analyzed glance parameters (i.e., total and mean glance durations). The HUD display position led to consistently longer eyes-on display times than the center console position. These findings corroborate previous studies on manual driving (e.g., Ecker, 2013; Hada, 1994). Furthermore, mean glance durations to the center console remained below three seconds in the current study, which is also consistent to previous findings across manual driving scenarios (Wikman, Niemenen, & Summalaa., 1998). The results of the current study might indicate participants’ awareness of potentially missing critical events (Tijerina, 2000) when glancing away from the road (Vollrath & Krems, 2011) during partially automated driving even when engaged in a secondary task. Moreover, the glance durations towards the HUD are potentially overestimated due to the participants possibly observing the driving environment peripherally or by looking straight through the display. Despite the longer eyes-on display time for the HUD, its closer proximity to the driving environment might enable a faster identification of and reaction to critical situations (e.g., caused by system failures) due to the reduced effort needed to shift attention between the AOIs (Wickens et al., 2003). Further research is needed to examine how longer eyes-on display times influence driver’s ability to monitor the system and instantly react to critical situations. Therefore, reaction times to and performance during critical events should be considered in further research.

ACKNOWLEDGEMENTS

This research was funded and supported by the BMW Group, Germany. All opinions expressed in this paper are those of the authors and not necessarily those of the BMW Group.
Effects of secondary tasks and display position on glance behavior during partially automated driving.

REFERENCES


Effects of secondary tasks and display position on glance behavior during partially automated driving.


ESRA: E-Survey of Road Users’ Attitudes – Analysis of Safety Indicators and Predictors of Distracted Driving Behaviour

Veerle Ross, Hasselt University, School for Transportation Sciences, Transportation Research Institute (IMOB), Belgium, veerle.ross@uhasselt.be, Kris Brijs, Hasselt University, School for Transportation Sciences, Transportation Research Institute (IMOB), Belgium, kris.brijs@uhasselt.be, Katrien Torfs, Knowledge Centre Road Safety, Vias institute, Belgium, Katrien.Torfs@vias.be, Muhammad Wisal Khattak, Ghent University, Civil Engineering & Hasselt University, School for Transportation Sciences, Transportation Research Institute (IMOB), Belgium, Muhammad.Khattak@ugent.be, Uta Meesmann, Knowledge Centre Road Safety, Vias institute & University of Liège, Urban & Environmental Engineering Department (LEMA), Belgium, Uta.Meesmann@vias.be

ABSTRACT

Background. Driving is a complex, goal-directed task with high perceptual, cognitive, and motor demands. Especially cell-phones and in-vehicle technologies increase driving complexity since they create dual-tasking situations behind the wheel. Distracted driving implies that attention, required for safe driving, diverts toward another competing activity, which may cause failures in hazard detection, information processing, vehicle control, etc. Objectives. The current study investigated distracted driving related to mobile-phone use. It aimed to identify a priority problem group and model the determinants of their distracted driving behaviour. Method. The data was collected by the ESRA project (E-Survey of Road Users’ Attitudes; Torfs et al., 2016; Trigoso et al., 2016). Data was collected in 2015 and 2016, using an internet based self-administered cross-sectional questionnaire from a sample of adults in 25 countries (e.g., Austria, Italy, Belgium, etc.). The total number of reported respondents was 26,643, of which 15,642 were regular car drivers. Four target behaviours related to the mobile-phone use while driving were examined: (1) handheld use (2) hands-free use (3) read text message or email (4) send text message or email. Results. “Young male drivers” were the identified problem group. Hands-free-phone use was the most prevalent mobile-phone use behaviour while driving. Relevant predictors were (specific terms between brackets): gender (males), education (middle aged drivers), higher exposure (young and middle aged), higher acceptability (low for hands-free use), attitudes (young), and support for traffic safety policy measures (old). These results can inspire policy-makers and legislation for priority-setting regarding distracted driving, especially for “young male drivers”.

1 INTRODUCTION

Driving a vehicle is complex, demanding task requiring full attention of the driver. Any distraction from within or outside a vehicle may impose a threat to the safety of the driver, occupants of the vehicle, or other road users. Distracted driving has been acknowledged as a large and growing road safety problem by the World Health Organization (WHO, 2015). In 2015, 85% of Americans judged driver distraction as the most dangerous threat to road safety, and they were more concerned about it compared to aggressive driving and drunk driving (AAA Foundation for Traffic Safety, 2016). The substantial growth of mobile-phone use while driving, and the increased representation of such drivers in road crashes, also attracted the attention of researchers (Dragutinovic & Twisk, 2005). Mobile-phone use while driving results in all types of distraction, i.e., visual,
physical, auditory, and cognitive. Generally, physical and visual distractions are linked with handheld mobile-phone use, while cognitive and auditory distractions are related to both hand-held and hands-free use. Whether it is hand-held or hands-free, distraction induced by mobile-phone use while driving causes a loss in driving performance, and thus should be avoided (Redelmeier & Tibshirani, 1997; Ross et al., 2014). The current study investigated distracted driving behaviour related to mobile-phone use while driving for data collected during the ESRA project (E-Survey of Road Users’ Attitudes (Waves 1+2); Torfs et al., 2016; Trigoso et al., 2016), which queried adults in 25 countries (e.g., Austria, Italy, Belgium, etc.). The objective is to identify a problem group among the participants and model determinants of distracted driving behaviour.

2 METHODS

2.1 The ESRA Project

ESRA is a joint initiative of about 38 countries around the world, aiming to collect a national level comparable data of the road users’ attitudes, opinions, and behaviour towards the road risks from active adult drivers. The ESRA-survey assumes fully comparable results among all the participating countries by ensuring a uniform method of sampling, using the same questionnaire, and a uniform programming technique for recording the responses from the participants. The current study analysed the data collected in 2016 from 25 countries during Waves 1+2 of ESRA-1 from 26,643 respondents of which 15642 were regular car drivers. Respondents were aged between 17 and 115 years (M= 44.73 years, SD= 15.11 years).

2.1.1 Questionnaire

A web-based self-administered questionnaire (SAQ) was used for the data collection. The questions included cover the socio-demographic information, the mobility and exposure to the traffic, road safety in general, acceptability of unsafe traffic behaviour, support for road safety policy measures, self-declared (unsafe) behaviour, attitude towards (unsafe) behaviour, subjective safety and risk perception, behaviour of other road users, involvement in road crashes and enforcement.

2.1.2 Measures

Four target behaviours related to the mobile phone use while driving in the ESRA questionnaire include (1) using a handheld mobile phone for calls (2) using a hands-free mobile for calls (3) using a mobile phone to read a text message or email (4) using a mobile phone to send a text message or email. Our target behaviours and other relevant questionnaire sections are: socio-demographic information, acceptability of behaviour, support for the road safety measures, self-declared behaviour, attitude towards (unsafe) traffic behaviour, and subjective safety and risk perception.

2.2 Data Analysis

The analysis used the characteristics of the drivers to predict the likelihood of distracted driving. A binary logistic regression was used to predict the likelihood that the driver will engage in the behaviours related to the mobile phone use while driving. Initially, an interlaced descriptive analysis determined the problem group. The association between the dependent variable and the independent (psychological) variables was assessed by
bivariate correlations. The variables used by the current study include the background information and the psychological variables associated with the drivers. Gender, age, educational level and the frequency of drivers were the background information used. The psychological variables comprised of the acceptability of the unsafe behaviours, the attitude towards unsafe driving behaviours and the support shown by the study participants to the traffic safety policy measures. The self-declared behaviour related to the use of mobile phone while driving in the last 12 months was the dependent variable. A summary of the results is provided in the following section.

3 RESULTS

3.1 Three-way Cross Tabulation (gender*age-group*self-reported behaviour)

Plotting percentage of the respondents using mobile phone while driving for different age-groups and genders in a three-way cross tabulation analysis gave an early warning of the problem road users’ group. The bar charts revealed that more male than female drivers have reported to talk on handheld and hands-free mobile phone while driving. All three age-groups have shown similar behaviour of male drivers when compared with female drivers. Similarly, greater percentage of male drivers than female drivers have used mobile phones for reading and sending emails and text messages while driving during last year. Drivers aged 18-24y have used a mobile phone for talking (handheld) and for reading and sending text messages and emails more than middle-aged and old-aged drivers (25-64y and 65 plus years respectively). However, similar percentage (60%) of young and middle-aged drivers reported using hands-free mobile phone while driving. All these results were validated by a chi square test of significance. The results are reported in figure 1.
3.2 Binary Logistic Regression Models for Talking on Handheld and Hands-Free Mobile Phone While Driving

Male drivers are more likely to report mobile phone use while driving than their female counterparts, i.e., lower odds ratio for females to perform behaviours related to mobile phone while driving (handheld: 0.78, hands-free: 0.75, read and sending text message/email: 0.86 both, where $p<0.001$). Age also has a significant effect as it decreases the likelihood of (a) handheld use for middle-age drivers (Odds ratio: 0.90, $p<0.01$) and for old-age drivers (Odds ratio: 0.44, $p<0.01$), (b) hands-free use for old-age drivers (Odds ratio: 0.55), (c) reading a text message/email for old-age drivers (odds ratio: 0.14) and (d) sending a text message in case of both middle-age (odds ratio: 0.53) and old-age drivers (odds ratio: 0.07). For all these results, $p<0.001$. Results further suggest a strong negative effect of the age on reading and sending text message/email than on the hands-free and handled mobile use while driving (Table 1). Higher educational levels increase the likelihood of using mobile phone by 1.23 for handheld, 1.54 for hands-free, 1.53 for reading and 1.51 for sending text message/email while driving for Bachelor’s degree holders and 1.25 for handheld, 1.43 for hands-free, 1.57 for both reading and sending text message/email while driving for Master and above degree holders (for all results, $p <0.001$). Frequent drivers are more likely to use a mobile phone while driving. E.g., the odds of a person who drives four days or more per week are 3.29 times higher to talk on handheld mobile than the one who drives only a few days per year. Similar results were observed for hands-free (Odds ratio: 3.66), reading a text/email (Odds ratio: 4.06) and sending a text and email (Odds ratio: 3.39) for drivers with four or more days of driving per week. For respondents driving 1-3 days per week or only a few days per month, the odds ratio (1.45 and 2.17 respectively) was relatively lower than the more frequent group (four or higher number of driving days, Odds ratio: 3.29) but higher than those who travel only few days a year. An increase in the likelihood for the hands-free use and reading and sending text messages and email was also observed (Table 1).

The acceptability of using a handheld and a hands-free mobile phone while driving increase the likelihood of engaging in these behaviours by 2.81 times ($p<0.001$) and 2.08 times ($p<0.001$) as compared to drivers who do
not accept or who have neutral opinions. Analysis of the acceptability of typing a text message or email and checking/updating social media showed a greater increase in the likelihood to read (Odds ratio: 4.33 and 3.36) and send text messages/email (odds ratio: 4.63 and 4.06). Attitude towards the unsafe behaviours, (i) attention to the traffic decreases when talking on handheld mobile phone and (ii) talking on handheld increase the risk of a crash, both decrease the odds of performing a handheld call while driving by 0.58 and 0.41 times. Drivers who believe that almost all the car drivers occasionally talk on handheld phone while driving are 2.35 times more likely to involve in such calls.

Support for policy measures has shown a decrease in using a mobile phone while driving. E.g., drivers supporting zero tolerance for using mobile phone while driving are 50% (Odds ratio: 0.50, P<0.001) and 53% (Odds ratio: 0.47, P<0.001) less likely to talk on handheld and hands-free mobile phone while driving. Similarly, support for the zero tolerance also decreased likelihood of reading (odds ratio = 0.41, p<0.001) and sending (odds ratio = 0.47, p<0.001) a text message or email while driving.

Table 1 - Binary logistic regression model for using a mobile phone while driving. Notes: (a) unweighted sample used (b) In Slovenia, the question ‘talks on a hand-held mobile phone’ refers do not limit it to hand-held mobile phone use only but it denotes all types of talking on the mobile phone while driving (c) *** p<0.001, ** p<0.01, * p<0.1

<table>
<thead>
<tr>
<th>Variables (Ref)</th>
<th>Handheld talking</th>
<th>Hands-free talking</th>
<th>Read texts/emails</th>
<th>Send texts/emails</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender (Ref: male)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.78***</td>
<td>0.75***</td>
<td>0.86***</td>
<td>0.86***</td>
</tr>
<tr>
<td><strong>Age-group (Ref: 18-24)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25-64</td>
<td>0.90*</td>
<td></td>
<td></td>
<td>0.53***</td>
</tr>
<tr>
<td>65+</td>
<td>0.44***</td>
<td>0.55***</td>
<td>0.14***</td>
<td>0.07***</td>
</tr>
<tr>
<td><strong>Educational Qualification (Ref: Primary education/none)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor’s degree or similar</td>
<td>1.23***</td>
<td>1.54***</td>
<td>1.53***</td>
<td>1.51***</td>
</tr>
<tr>
<td>Master’s degree or higher</td>
<td>1.25***</td>
<td>1.43***</td>
<td>1.57***</td>
<td>1.57***</td>
</tr>
<tr>
<td><strong>Frequency of driving (Ref: A few days a year)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A few days a month</td>
<td>1.45***</td>
<td>1.51***</td>
<td>1.80***</td>
<td>1.59***</td>
</tr>
<tr>
<td>1 to 3 days a week</td>
<td>2.47***</td>
<td>2.29***</td>
<td>2.66***</td>
<td>2.31***</td>
</tr>
<tr>
<td>At least four days a week</td>
<td>3.29***</td>
<td>3.66***</td>
<td>4.06***</td>
<td>3.39***</td>
</tr>
<tr>
<td><strong>Personal acceptability (Ref: unacceptable/neutral)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talk on a hand-held phone (acceptable)</td>
<td>2.81***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talk on hands-free phones (acceptable)</td>
<td></td>
<td>2.08***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type text messages or e-mails (acceptable)</td>
<td></td>
<td></td>
<td>4.33***</td>
<td>4.63***</td>
</tr>
</tbody>
</table>
## Social media (e.g., Facebook, twitter) (acceptable)

<table>
<thead>
<tr>
<th></th>
<th>Social media (e.g., Facebook, twitter)</th>
<th>3.36***</th>
<th>4.06***</th>
</tr>
</thead>
</table>

### Attitude towards unsafe traffic behaviour (Ref: disagree/neutral)

<table>
<thead>
<tr>
<th>Activity</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attention decreases when talking on a hand-held phone (agree)</td>
<td>0.58*** 0.46***</td>
</tr>
<tr>
<td>Almost all drivers occasionally talk on a hand-held phone (agree)</td>
<td>2.35***</td>
</tr>
<tr>
<td>People talking on a hand-held phone have a higher risk of an accident (agree)</td>
<td>0.41***</td>
</tr>
</tbody>
</table>

### Support for road safety policy measure (Ref: oppose/no opinion)

<table>
<thead>
<tr>
<th>Support for road safety policy measure (Ref: oppose/no opinion)</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero tolerance for any phone (hand-held or hands-free) for all drivers (Support)</td>
<td>0.50*** 0.47*** 0.41*** 0.47***</td>
</tr>
</tbody>
</table>

## 3.3 Binary Logistic Models for Young Drivers

The odds of using a mobile phone while driving for the young female drivers when compared with the young male drivers are reduced (handheld talking; odds ratio: 0.63, hands-free talking; odds ratio: 0.77, read a text message/email; odds ratio: 0.76). Education level is not a predictor of young drivers’ behaviour related to mobile phone use while driving except for reading the text messages which shows a significant decrease in the likelihood for the drivers with Master degree (odds ratio: 0.33, p<0.01). The frequency of driving and the odds of using a mobile phone while driving have positive relationship. The likelihood of using a handheld mobile phone to talk increases by 3.14 times, hands-free by 5.53 times, reading a text message/email by 4.77 times and sending a text message by 3.36 times for those who drive at least 4 days a week relative to those who drive only few days a year (for all odds ratio, p<0.001). Greater increase in the likelihood is observed for the hands-free and reading a text message/email (odds ratio: 5.53 and odds ratio: 4.77, respectively). The personal acceptability of using a mobile phone and the likelihood to engage in such behaviours while driving has shown a positive relation in the developed models. E.g., drivers accepting to talk on handheld and hands-free mobile phone while driving are 3.95 and 1.58 times more likely to perform these behaviours than those who do not accept. Similarly, young drivers accepting to type a text message or email and check or update a social media account are 5.40 times and 3.36 times more likely to read a text message and 3.33 times and 3.14 time to send a text message, respectively. Attitude towards unsafe behaviour is only significant for handheld mobile phone use while driving in our target age-group. People thinking that mobile phone use while behind the wheel decrease their attention to the traffic are 0.37 times (odds ratio: 0.63, p<0.01). The odds are increased by 2.57 times for drivers believing that almost all the drivers occasionally talk on handheld mobile phone while driving. When a driver has a highly negative attitude towards the mobile phone use while driving, e.g., it results in increased risk of getting an accident, the odds for these drivers are reduced by 0.62 times (odds ratio: 0.38, p<0.001). Drivers with attitude “My attention to the traffic decreases when talking on a hand-free mobile phone while driving” have 0.52 times odds of talking using hands-free mobile while driving. As expected, the support for zero tolerance about using a mobile phone for driving decrease the likelihood by 48% (odds ratio: 0.50*** 0.47*** 0.41*** 0.47***).
ESRA: E-Survey of Road Users Attitudes – Analysis of Safety Indicators and Predictors of Distracted Driving Behaviour

0.52, p<0.001), 43% (odds ratio: 0.57, p<0.001), 56% (odds ratio: 0.44, p<0.001) and 48% (odds ratio: 0.52, p<0.001), for handheld, hands-free, and read and send text message using mobile phone while driving, respectively.

Table 2 - Binary logistic regression models for age-group (18-24 years) using a mobile phone while driving.

Notes: (a) unweighted sample used (b) In Slovenia, the question ‘talks on a hand-held mobile phone’ refers do not limit it to hand-held mobile phone use only but it denotes all types of talking on the mobile phone while driving (c) *** p<0.001, ** p<0.01, * p<0.1

<table>
<thead>
<tr>
<th>Variables (Ref)</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Handheld phone</td>
</tr>
<tr>
<td>Gender (Ref: male)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.63***</td>
</tr>
<tr>
<td>Education (Ref: Primary education/none)</td>
<td></td>
</tr>
<tr>
<td>Secondary education</td>
<td></td>
</tr>
<tr>
<td>Bachelor’s degree or similar</td>
<td></td>
</tr>
<tr>
<td>Master’s degree or higher</td>
<td></td>
</tr>
<tr>
<td>Frequency of driving (Ref: A few days a year)</td>
<td></td>
</tr>
<tr>
<td>A few days a month</td>
<td>0.89</td>
</tr>
<tr>
<td>1 to 3 days a week</td>
<td>1.99*</td>
</tr>
<tr>
<td>At least four days a week</td>
<td>3.14***</td>
</tr>
<tr>
<td>Personal acceptability (Ref: unacceptable/neutral)</td>
<td></td>
</tr>
<tr>
<td>Talk on hand-held phone (acceptable)</td>
<td>3.95***</td>
</tr>
<tr>
<td>Talk on hands-free phone (acceptable)</td>
<td></td>
</tr>
<tr>
<td>Type texts/e-mails (acceptable)</td>
<td></td>
</tr>
<tr>
<td>Social media (e.g.: Facebook, twitter)</td>
<td></td>
</tr>
<tr>
<td>(acceptable)</td>
<td></td>
</tr>
<tr>
<td>Attitude towards unsafe traffic behaviour (Ref: disagree/neutral)</td>
<td></td>
</tr>
<tr>
<td>Attention decreases when talking on a hand-held phone (agree)</td>
<td>0.63**</td>
</tr>
<tr>
<td>Almost all drivers occasionally talk on a hand-held phone (agree)</td>
<td>2.57***</td>
</tr>
<tr>
<td>People talking on a hand-held phone have a higher risk of an accident (agree)</td>
<td>0.38***</td>
</tr>
<tr>
<td>Attention decreases when talking on a hand-free phone (agree)</td>
<td></td>
</tr>
<tr>
<td>Support for road safety policy measure (Ref: oppose/no opinion)</td>
<td></td>
</tr>
</tbody>
</table>

7
4 DISCUSSION AND IMPLICATIONS

The current study found support for the role of psychological factors in predicting the distracted driving related to the mobile phone use. The study also found that the contribution of the background information was somewhat unstable across the behaviours and this effect was more obvious when the variable age-group was controlled in the logistic regression. This fluctuating role of the background information in predicting odds of using a mobile while driving was also noticed when different age groups were compared. Other findings include that talking on the mobile phone while driving was more common than texting or emailing. Also, the study found that the participants are aware of the hazards associated with the distracted driving.

4.1 Role of the Background Information

Young drivers are more likely to use a mobile phone while driving than the middle-age drivers and the old age drivers. This result agreed to the findings of Kass et al. (2007) and Li et al. (2014) which has informed that with increase in the age, the self-reported distracted driving behaviour decreases. The current study found that the male drivers are more likely to use mobile phones while driving than the female drivers and this finding was common in all age-groups and for all the behaviours studied. Zhou et al. (2009) have also reported similar findings. Billieux et al. (2008) has contradictory findings for the behaviour related to sending text messages where females were found to send more messages. Educational level was a significant determinant for sending text messages in the old age drivers. For the middle age drivers, educational level was a significant predictor of the odds for all the studied mobile phone use behaviours. This finding can be linked to the results of Li et al. (2014). Li et al. (2014) found higher frequency of self-reported distracted driving behaviour with the higher earning. They argued that since income and educational level of the drivers are correlated, there is a chance that education had an indirect effect on the self-reported behaviour. Our analysis further revealed that the odds of using a mobile phone while driving are higher for those who drive frequently. An exception was the lack of a significant association between the self-reported behaviour and the frequency of driving in case of drivers aged 65 plus years. One explanation could be that since young- and middle-age people drive more frequently, the old-age people who drive frequent were reduced to a very small number in the analysis, making the analysis technique unsuitable.

4.2 Efficacy of Psychological Factors in Predicting Distracted Driving

The acceptability of the distracted driving was found to be positively associated with the self-reported behaviour of the participants. Li et al. (2014) also found that the drivers accepting more easily the distracted driving behaviour (all mobile phone use related) were more likely to report engaging in the distracted driving. The respondents’ attitude towards the distracted driving and support for traffic safety policy were other useful psychological predictors. This finding followed the results of Li et al. (2014) which suggested that people who were engaged in the distracted driving were less likely to see this behaviour as a serious safety concern. The differences in the capability of the variables to predict odds ratio to involve in the distracted driving varied
across different age groups. By dividing the study sample into three new age-group, a loss of balance was induced into the number of participants in each group. Given the huge size of the middle-age drivers group (25-64y), the number of participant in that group was almost eight to nine times more than the other age-groups. Trigoso et al. (2016) has induced an artificial balance between the sample sizes of different age-groups by using weights and almost equal age-group sizes (i.e. 18-34y, 35-54, and 55+).

4.3 Implications

Results of the current study can inspire policy-makers and legislation for priority-setting regarding distracted driving, especially for “young male drivers”. The results showed that attitude towards the unsafe behaviour can be associated with the likelihood of engaging in the distracted driving. The attitude of the drivers towards the (unsafe) traffic behaviour can be included in behaviour change campaigns. Attitudes like engaging in the mobile-phone conversation while driving ‘reduces attention to traffic’ and ‘increases the risk of getting into an accident’ can be important campaign targets. These campaigns should be targeted especially to young male drivers, thus including very specific strategies.

REFERENCES


AUTOMATED VEHICLES: TRUST, ACCEPT AND MOVE ON?

Parallel session
Fostering Trust and Acceptance of a Collision Avoidance System through Retrospective Feedback

David R Large*, James Khan, Gary Burnett
Human Factors Research Group, University of Nottingham, UK, *david.r.large@nottingham.ac.uk

ABSTRACT
A simulator study explored the effects of providing retrospective feedback on drivers’ acceptance of a collision avoidance system (CAS) following a false activation. Sixteen experienced drivers undertook two drives, each lasting approximately 20 minutes. During both drives, the CAS identified a rogue pedestrian and intervened by providing an audible warning closely following by emergency braking. Feedback was provided following one of the activations (with the order counterbalanced between participants), in the form of a detailed storyboard ‘playback’ depicting the system’s analysis of the situation. Subjective ratings of trust, confidence, annoyance and desirability, revealed no differences overall between conditions (i.e. with and without post-event feedback). However, there was a tendency for drivers to trust the system more if feedback was provided during their first drive, whereas drivers who were provided feedback during their second drive indicated higher levels of confidence in the system and found it less annoying. There was also a significant rise in the number of drivers who detected the potential pedestrian hazard prior to system activation during their second drive. Results suggests benefits associated with the provision of retrospective feedback, but effects may have been influenced by the experimental design, which exposed participants to similar pedestrian hazard events in consecutive drives. Future investigations, which should continue to explore techniques to enhance trust and acceptance of active safety systems, should therefore adopt a between-subjects design to isolate effects.

Keywords: trust, acceptance, collision avoidance system, retrospective feedback.

1 INTRODUCTION
Accidents involving vulnerable road users (VRUs) remain a major concern for road safety, accounting for almost 40% of road fatalities in Europe, and almost 50% worldwide (WHO, 2015). Pedestrians are one of the most vulnerable road user groups, both in terms of the likelihood of being involved in a near-miss or collision, and the potential ramifications should an incident occur (Clifton et al., 2009). Active safety systems and collision avoidance systems (CAS), such as Pedestrian and Cyclist Detection Systems with Emergency Braking (PCDS+EBR), have the potential to mitigate the risk to VRUs by warning drivers of an impending collision and/or taking evasive action autonomously by braking, steering or both, if a collision becomes imminent or if the driver fails to respond. Moreover, evidence suggest that PCDS+EBR systems have the greatest potential to improve the safety of VRUs, with data indicating a reduction of 7.5% on all road fatalities and 5.8% on all road injuries, representing an estimated 2,100 fatalities and over 62,900 injuries saved per year in the EU-28 based on 2015 accident trends (Silla et al., 2015).

However, current limitations in detection technology and algorithms, combined with the inherent difficulty of predicting pedestrians’ behaviour and intentions mean that even state-of-the-art PCDS can be prone to high
numbers of false alarms and/or activations (Zhang et al., 2016). Moreover, as the technology moves from ‘collision mitigation’ (with the aim of reducing vehicle speed/kinetic energy to make the impact more survivable) to ‘collision avoidance’ (aiming to avoid collisions completely), the number of false alarms and/or activations is naturally expected to increase as the margins of error are much smaller. The frequency and occurrence of false alarms and activations will inevitably influence drivers’ attitudes towards the system, potentially encouraging them to neglect it, find creative ways to bypass it, or deactivate it completely (Parasuraman and Riley, 1997).

Given that estimated benefits assume full market penetration and complete use (Silla et al., 2015), drivers’ acceptance of the technology is therefore important.

In a driving context, acceptance has been described as “the degree to which an individual incorporates the system in his/her driving, or if the system is not available, intends to use it” (Adell, 2009). The determinants of acceptance are therefore complex and derive from a multitude of factors, including trust, the driver’s experience of interaction with technology, their understanding of system limits, and the context in which it is implemented. Thus, factors, such as the number and frequency of false alarms, are likely to play a significant role in shaping drivers’ trust and acceptance. Moreover, human behaviour is not primarily determined by objective factors, but also by subjective perceptions (Ghazizadeh et al. 2012). This means that acceptance is based on individual attitudes, expectations and experience as well as the subjective evaluation of expected benefits (Schade and Baum, 2007). It has even been suggested that the degree of technological innovation has a lesser effect on acceptance than personal experience (Ausserer and Risser, 2005).

Providing feedback to the user has been shown to have a significant positive effect on the development of trust and acceptance of technology, as it allows the individual to judge system expertise (Miller et al., 2014). From a system-design perspective, feedback can be considered as the information available to the operator regarding the state of the joint human-machine system, and can be provided to enhance immediate performance or induce behavioural change (Donmez, Boyle and Lee, 2008). In a driving context, immediate performance feedback is often inherent within the driving task itself (e.g. headway, lane position), and enables drivers to calibrate and modify their performance or behaviour as necessary, for example, by moderating their engagement in distracting secondary activities (Donmez et al., 2007). Feedback is also often provided regarding the status of on-board safety systems via various interfaces. However, this may be limited to binary states of ‘system activated’ or ‘system idle’. The concern is that a false activation (for example, where the system mistakenly predicts that a pedestrian will enter the vehicle’s trajectory and then applies emergency braking), may startle drivers and leave them perplexed regarding the system’s actions, particularly if they were unaware of any potential risk; this will likely detriment their trust and acceptance of the technology.

Providing more detailed feedback regarding the behaviour or intentions of a PCDS system will likely increase trust and acceptance, but may interfere with primary task performance and distract drivers if provided in real-time (Arroyo et al., 2006). Moreover, the limited time that can be allocated to concurrent feedback makes it impossible to provide detailed information regarding the event that triggered the warning or system activation. As a consequence, concurrent feedback may not convey the information necessary to explain or guide behaviour. In contrast, providing feedback retrospectively (e.g. when the driver has stopped or at the end of a journey) enables the sharing of richer information (Donmez, Boyle and Lee, 2008). For discretionary systems (e.g. systems designed...
Fostering Trust and Acceptance of a Collision Avoidance System through Retrospective Feedback

to improve driving performance and behaviour), this can support the memory of critical incidents and help drivers understand how their behaviour may have contributed to these (Donmez, Boyle and Lee, 2008). It is hypothesized that providing retrospective feedback to drivers regarding the intentions and behaviour of an active safety system, particularly following a false activation, will increase their overall acceptance and trust of the technology. The current investigation explores this using self-reported ratings of trust, confidence, annoyance and desirability, and drivers’ visual behaviour.

2 METHOD

Sixteen experienced drivers (10 male, 6 female, age range 23-56, mean age 28.93, mean time with license 10.3 years) undertook two drives, each lasting approximately 20 minutes, in a medium-fidelity, fixed-base driving simulator. The simulator comprises an Audi TT car located within a curved screen, providing 270° forward and side image of the driving scene via three overhead HD projectors. Rear view mirror images are captured digitally and relayed to two 7-inch LCD screens, located to replicate the side mirrors, and a 55-inch curved HD LED television positioned behind the vehicle and visible using the existing interior rear-view mirror. A Thrustmaster 500RS force feedback wheel and pedal set are faithfully integrated with the existing Audi controls. The driving scenario was created using STISIM Drive (v3) software to replicate a mixed driving environment, including residential, urban/town and rural components. Participants wore SMI eye-tracking glasses to capture their visual behaviour, and received a £10 (GBP) shopping voucher as compensation for their time.

Towards the end of each drive, as the vehicle passed through a busy urban environment, a pedestrian walked towards the roadside, as if intending to cross (Figure 1). The PCDS identified the rogue pedestrian (walking at a constant speed of 1.0 m/sec) as a potential hazard, providing an audible warning to alert the driver at a time-to-collision (TTC) of 1.2s, and initiated an emergency braking manoeuvre when the TTC was 0.5s, ultimately bringing the car to a stop at a clearance distance of 3.8m. The pedestrian remained in full view of driver throughout the confrontation, reflecting the most common accident use-case (PROSPECT, 2016). Despite the system’s analysis of the situation, the pedestrian actually stopped at the road-side. Thus, the event was expected to be perceived by the driver as a ‘false positive’ intervention.

Following the ‘false’ activation, the driver was asked to pull over to the side of the road when it was safe to do so. During one of the two drives (which were counterbalanced), the driver was then provided with retrospective feedback depicting a detailed storyboard ‘playback’ of the system’s analysis of the situation (Figure 2) presented on a Microsoft Surface tablet computer mounted in the centre console of the vehicle. No feedback was provided during the other drive. After each drive, ratings of trust were obtained using the trust in automation questionnaire (Gold et al., 2015, adapted from Jian et al., 2000). This was supplemented by additional bespoke items exploring concepts such as participants’ confidence in the system, their annoyance with it, and whether they would choose to have the system in their own vehicle, with responses invited using Likert scales.
3 RESULTS

Cumulative ratings of trust, confidence and annoyance were calculated. Overall, these revealed no significant differences between conditions (i.e. with and without post-event feedback), but indicated lower trust associated with the second drive ($t(15)=2.83, p=.013$), suggesting a potential order effect (for drive 1, mean = 3.63; drive 2 = 2.98). Interrogating these data further, however, it was evident that there was a tendency for drivers to trust the system more if feedback was provided during their first drive. Conversely, drivers who were provided feedback during their second drive indicated lower levels of trust, but higher levels of confidence in the system (Figure 3) and found it less annoying, although ratings of annoyance were generally low throughout. Drivers also indicated that they were more likely to want the system in their own car based on their experience of feedback during their second drive. Eye-tracking analysis revealed a significant rise in the number of drivers who detected the potential pedestrian hazard prior to system activation during their second drive (25% in drive 1 and 81% in drive 2) ($t(15)=4.39, p = .001$), suggesting increased vigilance, although it is unclear whether this was a behavioural change inspired by the provision of retrospective feedback during drive 1, or an experimental experience effect (i.e. expectation following repeated exposures to similar pedestrian hazard events in consecutive drives).
Fostering Trust and Acceptance of a Collision Avoidance System through Retrospective Feedback

4 DISCUSSION

The study explored the effects of providing detailed post-event feedback on drivers’ acceptance of a PCDS following a false activation. While the results did not fully support our hypothesis that providing retrospective feedback would increase drivers’ overall acceptance and trust of the technology, there are some encouraging findings, which are significant in light of existing research and can be used to inform future studies. For example, previous research has shown that exposure to the first false alarm has the strongest negative effect on trust ratings (the so-called ‘first failure effect’; Rovira et al., 2007). Therefore, providing feedback during the first exposure (i.e. the first drive in our study) to counteract this effect, might be expected to elicit higher trust ratings, and this tends to be supported by our data. Nevertheless, some of the findings may be confounded by our experimental design, which adopted a within-subjects approach (due to timing and resource constraints), whereby each participant was exposed to both conditions (i.e. with and without feedback). Although the order of exposure was counterbalanced, it is feasible that exposure to the potential hazard and associated warning/activation during the first drive may have encouraged drivers to be more vigilant during the second drive. The fact that a greater proportion of drivers saw the pedestrian threat during the second drive, prior to system activation, tends to support this statement, although, it is unclear how this may have impacted their behaviour. For example, if drivers were already aware of a potential threat, they may have expected the activation, thereby increasing their confidence in the system when it acted appropriately, as the data suggest – although this may not have been in response to the feedback per se. Alternatively, the fact that drivers were already aware of the risk, may have meant that they felt that the intervention was unnecessary, and this could have reduced their confidence in the system or annoyed them (contrary to our findings).

Consequently, it is not possible to isolate effects, making it difficult to draw any robust conclusions. In addition, the relatively small sample (and subsequent between-subjects analysis) limits the statistical power and potential impact of the research. As such, further work is required. In particular, it is recommended that future investigations, which should continue to explore the benefits of providing retrospective feedback on trust and acceptance as the literature encourages, adopt a between-subjects design with a larger cohort of drivers.
5 ACKNOWLEDGEMENTS

The research received funding from the European Community’s Eighth Framework Program (Horizon2020) (agreement no. 634149), as part of the Proactive Safety for Pedestrians and Cyclists (PROSPECT) project.

REFERENCES


Silla, A., Rama, P., Leden, L., van Noort, M., de Kruijff, J., Bell, D., Morris, A., Hancox, G. and Scholliers, J., 2015. Are intelligent transport systems effective in improving the safety of vulnerable road users?


Zhang, S., Benenson, R., Omran, M., Hosang, J. and Schiele, B., 2016. How far are we from solving pedestrian
Fostering Trust and Acceptance of a Collision Avoidance System through Retrospective Feedback

Gaze Behaviour as a Measure of Trust in Automated Vehicles

Francesco Walker, University of Twente, The Netherlands, f.walker@utwente.nl, Willem Verwey, University of Twente, The Netherlands, Marieke Martens, University of Twente, The Netherlands

ABSTRACT

Extensive evidence shows that drivers’ intention to use self-driving technology is strongly modulated by trust, and that the benefits promised by automated vehicles will not be achieved if users do not trust them. It follows that vehicles should be designed to achieve optimal trust. However, there is no consensus as to how this should be assessed. To date, most studies have relied on self-reports, an approach subject to problems. In the driving simulator study reported here, we used drivers’ gaze behaviour during simulated Highly Automated Driving to investigate whether this could provide a more objective measure. The results indicated a negative relationship between self-reported trust and monitoring behaviour: The higher their self-reported trust, the less participants monitored the road, and the more attention they paid to a non-driving related secondary task. These findings suggest that gaze behaviour in a secondary task situation provides a more objective indicator of driver trust than self-reports.

Keywords: trust in automation, eye movement behaviour, human factors, autonomous driving, automated vehicle, secondary task.

1 INTRODUCTION

Studies show that more than 90% of road accidents are due to human errors (e.g. NHTSA, 2008). Automated vehicles are expected to drastically reduce the number of accidents. They are also expected to increase travelling comfort, allowing users to engage in all sort of activities while the car takes care of driving, and giving back the freedom to travel to individuals who have lost the ability to drive due to age or disability (Payre, Cestac, & Delhomme, 2016; Urmson & Whittaker, 2008). Nevertheless, none of these benefits will be achieved if users do not trust the automated system.

Extensive evidence indicates that acceptance, perceived usefulness and intention to use self-driving technology are strongly mediated by trust (Choi & Ji, 2015; Ghazizadeh et al., 2012; Parasuraman & Riley, 1997). Importantly, trust evolves dynamically over time. This means that drivers’ intention to use the technology depends not just on their initial trust levels, but also on their experience with the automated system and on its perceived reliability (Körber, Baselar, & Bengler, 2018; Sauer, Chavaillaz, & Wastell, 2016; Lee & See, 2004).

To understand how trust impacts the use of Highly Automated Vehicles we need effective trust metrics. Up till now, most studies have relied on self-reports. However, questionnaires are not a continuous measure. This means they cannot capture real-time changes in user trust, and are hard to use outside an experimental context (Hergeth, Lorenz, Vilimek, & Krems, 2016). Analysis of drivers’ gaze behaviour could potentially provide a continuous and objective measure of trust.

To date, there have been few studies on the relationship between driver trust during Highly Automated Driving (HAD) and eye-movement behaviour. A few (e.g. Körber et al., 2018; Hergeth et al., 2016; Hellandin, Falkman, Riveiro, & Davidsson, 2013) suggest that participants with a high level of trust tend to monitor the road less. Others (e.g. Gold et al., 2015) have failed to find such a link. The aim of our study is to generate new evidence that can help to resolve the question whether eye movement behaviour may provide a reliable indicator for the trust of drivers in automated vehicles. Importantly, if drivers’ trust could be objectively measured in real-time, vehicle behaviour and display information could then be tuned accordingly.

In the present study, we investigated the influence of the vehicle reliability on drivers’ monitoring behaviour. Videos were used to simulate HAD. In the videos, the simulated vehicle performed longitudinal and lateral vehicle control, and applied the brakes when cyclists or pedestrians were crossing the road. Participants were asked to pay attention to the road and perform a secondary task, but only if they trusted the way the system was handling the driving. We compared two groups of participants, where each participant was seated in the mock-up of the driving simulator. One group viewed videos of a car handling the driving task perfectly; a second group viewed videos of a car struggling with the driving task (i.e. it tended to drift towards the centre of the road and braked abruptly when approaching crossing pedestrians or cyclists). In line with Hergeth et al. (2016) and Korber et al.’s
Gaze Behaviour as a Measure of Trust in Automated Vehicles

(2018) results, we expected to find a negative relationship between trust and monitoring frequency: The less drivers trust the system, the more they view the road, and vice versa.

2 METHODS

2.1 Participants

Thirty participants, all students of the University of Twente, were selected for the experiment and participated in exchange for money (6 euro) or study credits. Six participants were excluded from the analysis: 5 due to the poor quality of their eye-tracking data (i.e. pupil detected in less than 79% of the frames), and 1 because the post interview showed that the participant had previous experience with a Level 2 (SAE, 2014) automated vehicle. The other participants reported no previous experience with automated vehicles. All participants had their driver’s licence for at least one year and reported normal vision. None wore glasses, and none reported that they commonly suffered from motion sickness. The 14 female and 10 male participants selected for the final analysis were all between 18 and 24 years of age (M = 20.46; SD = 1.414).

2.2 Videos – Pilot study

A Go-Pro Hero 4 Session camera, placed centrally on the hood of an Audi A4, was used to film 14 two-minute driving scenarios, encountered while driving in the surroundings of the University of Twente. We then performed a pilot study in the driving simulator of the University of Twente, in which 10 participants viewed the videos in a different random order. One participant was excluded due to motion sickness. The remaining participants were between 19 and 28 years of age (M = 21.7; SD = 2.3). They were asked to imagine being in an autonomously driving vehicle. After each video they were asked to verbally rate the vehicle’s performance on a 7-point Likert scale. Of the 14 videos, the 3 with the highest scores (i.e. M > 5.7) and the 3 with the lowest scores (i.e. M < 2.8) were used in the main experiment. Wilcoxon signed ranks tests showed that the scores of the 3 highest rated videos significantly differed from the 3 scores with the lowest ratings (all p’s = .005).

2.3 Tasks

Participants were assigned to one of two groups: The Perfect Vehicle group viewed forward looking videos recorded from a car handling the driving task perfectly (it always kept in lane, and braked comfortably in front of crossing cyclists or pedestrians). The Poor Vehicle group viewed videos of a car struggling with the driving task, with a tendency to drift towards the centre of the road, and to brake abruptly in front of crossing cyclists and pedestrians. Perfect Vehicle and Poor Vehicle participants were both confronted with one crossing cyclist or pedestrian per video. Although group assignment was random, the randomization procedure ensured the same distribution of participants by gender and initial trust level in each group.

Participants of both groups performed a non-driving-related secondary task (NDRT) during simulated driving. This was presented on a Dell Latitude 7370 laptop, and was programmed using Psychopy software (Peirce, 2009). As in Verwey (2000), the NDRT consisted of a simple addition task. In each trial, a target number, between 20 and 87, was presented in red at the centre of the screen. Below the target number, the screen showed an addition in the form <number>+7. As previously, the number always lay in the range from 20 to 87. Participants were asked to indicate whether the result of the addition was higher or lower than the target number. At the end of each trial, system feedback (the word “correct” for correct responses, the word “incorrect” for incorrect responses) was presented centrally on the laptop screen. Participants had no time limit to perform the task. Responses were recorded using the right and left arrows of a standard keyboard, which participants held on their lap for the entire duration of the experiment. The task is presented in Figure 1.

![Figure 1. Secondary addition task. Correct trial: By pressing the left arrow, the participant correctly indicated that 76+7 is lower than the target number “86.”](image-url)
To perform the NDRT participants allocated their attention away from the driving simulator screen, and could not use peripheral vision to watch the road while they were performing the secondary task. The time participants were watching the NDRT screen served as an indicator of trust.

2.4 Trust questionnaire

Participants’ trust in the simulated vehicle was also measured using a modified version of the Empirically Derived (ED) Trust Scale (Jian, Bisanz, & Drury, 2000). This uses a 7-point Likert scale (1 = totally disagree; 7 = totally agree), to indicate level of agreement with seven statements (based on Verberne, Ham, & Midden, 2012). Participants filled in the trust questionnaire before and after completing the experiment (see section “Procedure”). Assessing trust before the start of the experiment was important to ensure that participants in each group were equally distributed in terms of their initial trust.

2.5 Apparatus

2.5.1 Driving simulator

The driving simulator of the University of Twente consists of a mock-up equipped with steering wheel, pedals and indicators. Videos, displayed through Psychopy software (Peirce, 2009), are projected on a screen of 7.8 x 1.95 meters. The screen has a total resolution of 3072*768 pixels (~10ppi).

2.5.2 Mobile eye-tracker

Participants’ eye-movements were recorded using Tobii Pro Glasses 2. The head-mounted mobile eye-tracker weighs 45 grams and is equipped with four eye cameras. It tracks movements of both eyes, and uses an additional scene camera to track the external world. The glasses were connected to a 130 x 85 x 27 mm recording unit with a weight of 312 grams. The eye-tracker was wirelessly connected to a Dell tablet, running Windows 10 operating system and the Tobii Pro Glasses Controller software (Tobii Pro Glasses 2, 2018).

2.6 Procedure

Participants filled out a pre-measurement trust questionnaire online at home. Then, on the testing day, they were welcomed to the driving simulator room and told that they would view videos while sitting in the simulator mock-up. After filling in an informed consent form, the mobile eye-tracker was calibrated following the procedure recommended by Tobii (i.e. Tobii Pro Glasses 2, 2018). After calibration, participants sat down in the mock-up of the driving simulator. On screen instructions informed them that the experiment would be divided into two phases. Phase 1 would be a “trust development session”, in which they could develop a general idea of how the system worked by watching its behaviour. They were told that in this phase they should focus on the vehicle’s behaviour. In phase 2 the car would behave as in phase 1, though the scenarios would be different. The phase 1 video was then presented on screen. At the end of the video, participants were asked “How do you think that the car coped with the driving task?”. Responses were expressed out loud using a 7-point Likert scale, with 1 indicating “very badly” and 7 “very well”.

After participants’ responses had been recorded, instructions for phase 2, the “active session”, were displayed on screen: Participants were told to imagine they were in a self-driving car, and to perform the secondary task whenever they felt that the car was handling the driving safely. Once again, participants were reminded that the car would behave as in phase 1. Participants were given time to practice the secondary task before the start of the phase. After practice, two videos were presented. At the end of phase 2, participants were asked to rate how they thought the car had coped with the driving task. The order of the videos was counterbalanced across participants.

Participants were then asked to leave the mock-up, remove the eye-tracker, and fill in the post trust questionnaire. Although the items of the post-measurement were the same as those in the pre-measurement questionnaire, they were rephrased to refer to participants’ trust towards the vehicle they had just experienced, and not their general trust in self-driving cars. The experiment ended with a final questionnaire, in which participants were asked to provide information on their level of education and previous experience with automated vehicles. Responses were used to check that participants had no previous experience with automated vehicles, either as drivers or as passengers.
Gaze Behaviour as a Measure of Trust in Automated Vehicles

2.7 Analysis

Our analysis assessed participant eye-movement behaviour during phase 2 onto two regions of interest (ROI): 1) the road, i.e., the central section of the driving simulator screen (size 2.6 x 1.95 meters); 2) the laptop, on which the NDRT was presented. Specifically, we assessed fixation count (i.e. number of fixations made in each ROI) and fixation duration (i.e. total time spent viewing each ROI).

3 RESULTS

3.1 Monitoring behaviour and trust

Analysis of the results for the two groups showed that 65% (SD = .099) of the Perfect Vehicle group’s fixations were on the road, as opposed to 83% (SD = .084) for the Poor Vehicle group. A Mann Whitney U test revealed that the difference was highly significant (U = 12, \( p = .001 \)). Moreover, members of the Perfect Vehicle group spent 70% (SD = .117) of their time looking at the road, as opposed to 84% (SD = .113) for the Poor Vehicle group. Again, the difference between the two groups was highly significant (U = 24, \( p = .007 \)). These results are strong evidence that monitoring behaviour could be an effective measure of trust. See Figure 2.

![Figure 2. Participant monitoring behaviour. The Poor Vehicle group made more fixations (83% vs. 65%) and spent more time (84% vs. 70%) on the road compared to the Perfect Vehicle group. Error bars represent standard error of the means.](image)

To confirm the relationship between trust and monitoring behaviour, we used a Pearson correlation to correlate the two measures. Self-reported trust scores collected at the end of the study and pooled across the two groups correlated strongly both with fixation count (\( r = -.658, p < .001 \)) and with fixation duration (\( r = -.611, p = .001 \)): In brief, the higher the trust, the less participants fixated the road. See Figure 3.

![Figure 3. Correlation between self-reported post trust (1: low; 7: high) and monitoring behaviour. The higher the trust, the less participants fixated on the road. A) Fixations made on road; B) Time spent on road.](image)
3.2 Validation

Two independent annotators used Tobii Pro Lab software to classify the same set of fixations on the road and on the laptop. Interrater reliability was estimated using a Pearson correlation. Fixation durations and fixation counts computed from the fixations mapped by judge 1 correlated strongly with the fixations mapped by judge 2 (durations: $r = .998$, $p < .001$; counts, $r = .999$, $p < .001$).

To test whether the videos presented to the Perfect Vehicle group were associated with higher self-reported trust than those presented to the Poor Vehicle group, we analysed the data with a Mann-Whitney U test. As expected, the Perfect Vehicle group had higher scores. This was true both for the videos presented during phase 1 (Perfect Vehicle: $M = 6.2$, $SD = .632$; Poor Vehicle: $M = 4.93$, $SD = 1.328$; $U = 30.5$) and phase 2 (Perfect Vehicle: $M = 5.3$, $SD = .949$; Poor Vehicle: $M = 3.71$, $SD = 1.541$; $U = 29$), $p = .016$. These findings are further evidence that monitoring behaviour can indeed provide a valid measure of driver trust.

4 DISCUSSION

Extensive evidence shows that adoption of automated technology is strongly modulated by user trust. Up to now, however, driver trust in self-driving technology has usually been investigated through self-reports – a technique that is not continuous, and difficult to use in real-world scenarios. To better understand the link between driver trust and the behaviour of automated vehicles, we need trust measures that are more objective and easier to use. Here, we investigated the potential of mobile eye-tracking technology in a secondary task situation. Participants were divided into two groups. The Perfect Vehicle group viewed videos of a simulated Highly Automated Vehicle (SAE, 2014) which coped perfectly with the driving task, while the videos presented to the Poor Vehicle group showed a vehicle that tended to swerve towards the centre of the road and braked abruptly in front of crossing cyclists or pedestrians. Participants’ eye-movements to the road and the secondary task were recorded, and self-reported trust was measured before the start (for control purposes) and at the end of the experiment.

Comparison of the two measures showed that the higher the drivers’ self-reported trust scores collected at the end of study, the less time they spent viewing the road, and vice versa. These results confirm the negative relationship between trust and monitoring behaviour (i.e. fixation count and duration) during Highly Automated Driving (SAE, 2014), previously reported by Hergeth et al. (2016) and Korber et al. (2018).

The results also confirm the success of our manipulation. The Perfect Vehicle group made more fixations and spent more time on the secondary task compared to the Poor Vehicle group. Participants were instructed to perform the secondary task only when they thought that the automated system was handling the driving task safely. It appears, therefore, that the Perfect Vehicle group trusted their simulated vehicle more than the Poor Vehicle group. This was confirmed by participants’ answers to the question “How do you think that the car coped with the driving task?”. Both in phase 1 and phase 2, the Perfect Vehicle group reported higher scores than the Poor Vehicle group.

One limitation of our study is that participants who thought that the automated vehicle was not handling the driving task safely were instructed to keep viewing the road. In a real-world scenario, it is likely that drivers who felt uncomfortable with the car’s behaviour would disengage the automated system and take back manual control. Here, due to the video-based nature of the study, drivers could not disengage the automated system. It is also likely that participants in this simulation-based study did not feel as threatened by the vehicle’s behaviour as they would have felt in a real-world situation. Nonetheless, results clearly show different eye-movement behaviour between the two groups.

To summarize, our study represents a step forward in the development of more objective, non-invasive and continuous measurements of drivers’ trust in self-driving technology. Such measurements are particularly important in real-world scenarios, where they could be used to assess real-time changes in driver trust. If drivers over- or underestimate the capabilities of their vehicles, automated systems could react by modifying their behaviour or by providing them with additional information concerning the vehicles’ performance. As suggested by Hergeth et al. (2016), future research should further investigate the interactions between self-reported trust, monitoring behaviour and driver behaviour. This is a necessary step towards the use of gaze behaviour as an objective and continuous measurement of driver trust.
REFERENCES


Helldin, T., Falkman, G., Riveiro, M., & Davidsson, S. (2013, October). Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving. In Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 210-217). ACM.


Autonomous Vehicle Validation – are we just guessing or can we predict the future?

Steven Reed, Loughborough University, UK, s.g.reed@lboro.ac.uk, Laurie Brown, Loughborough University, UK

ABSTRACT

At present there exists no standardised, systematic or structured methodology for the validation of autonomous pod vehicles. The lack of process is compounded by minimal real-world collision or near miss data covering this vehicle type. This gap in knowledge may present a significant obstacle in autonomous pod vehicle development and consumer uptake from two sides; firstly, the scarcity of data for road safety practitioners to develop into appropriate test scenarios and secondly the lack of perceived transparency by the OEMs could form the perception that there is negligible vehicle development and manufacturer accountability, resulting in lack of trust from the end users.

To counter this knowledge gap this paper aims to provide more information on initial steps into the definition of a systematic reference dataset which reflects both autonomous pod use and the capabilities of the vehicles to be tested. Due to the lack of real world data in the short term, it will be necessary to develop vehicle validation scenarios for autonomous pods based entirely on non-autonomous collisions of comparable vehicle types. Although there are currently no specific well-tested frameworks to follow, the approach discussed applies proven methodological principals from the field of general product design which assumes that if the design envelope is set correctly then any product that meets this, no matter how outlandish, will be valid.

Keywords: Autonomous vehicles, autonomous pods, validation, test scenarios, accident data.

1 INTRODUCTION

In recent years there has been a rapid rise in the number of vehicles with autonomous features coming to market, and the race between manufacturers to get fully autonomous cars onto the road has driven quick advances in this technology with relatively little validation procedures compared to standard vehicles [1]. Additionally, these advances have led to the introduction of new vehicle types, as these partially-autonomous vehicles include not only traditional car designs (e.g. Tesla, Ford, BMW), but also smaller category vehicles such as the as-yet unclassified group of vehicles known as ‘pod’s (e.g. Gateway [2], Navya [3]).

At present there exists no standardised, systematic or structured methodology for the validation of these autonomous ‘pod’ vehicles. This is due, in part, to the small number of any types of vehicles with autonomous features that exist in the overall vehicle fleet, but also because useful information relating to any collision is typically very tightly regulated by the Original Equipment Manufacturer (OEM). For pods, the lack of process is compounded by minimal real-world collision or near miss data covering this vehicle type. This gap in knowledge may present a significant obstacle in autonomous pod vehicle development and consumer uptake from two sides; firstly, the scarcity of data for road safety practitioners to develop into appropriate test scenarios, and secondly the lack of perceived transparency by the OEMs could form the perception that there is a lack of
vehicle development and manufacturer accountability, resulting in lack of trust from the end users.

To counter this knowledge gap this paper aims to provide more information on initial steps into the definition of a systematic reference dataset which reflects both the intended use of autonomous pods and the capabilities of the vehicles to be tested. Due to the lack of real world data in the short term, it will be necessary to develop vehicle validation scenarios for autonomous pods based entirely on non- or partially-autonomous collisions of comparable vehicle types. Although there are currently no specific well-tested frameworks to follow, the approach discussed applies proven methodological principals from the field of general product design which assumes that if the design envelope is set correctly then any product, no matter how outlandish, that meets this will be valid.

This approach applied to collision data underpins the autonomous pod validation framework and likewise assumes that any scenario derived from the dataset to the proposed pod vehicle will provide close alignment to only valid real-world collisions for use in the test scenarios. This new approach is currently being developed through the Capri (Connected & Autonomous POD on-Road Implementation) project, an industry-led research and development project funded by Innovate UK [4], which aims to build and test the next generation of autonomous PODs as well as the systems and technologies that will allow the vehicles to navigate safely and seamlessly in both pedestrian and road environments.

2 DATA APPROACH

Currently scenario generation for autonomous vehicles is a rapidly growing area of research. There are typically two ways of doing this, firstly by generating scenarios from a theoretical basis, for example basing the testing program on a list of accident types, and secondly from a practical ‘beta-testing’ basis, for example, pressing the vehicles into service to see what they encounter [5].

Both of these approaches will most likely be incomplete and may ultimately fail to deliver the range of scenarios necessary to build consumer trust. By their very nature collisions are rare events but they also vary considerably based on a huge range of independent factors. Without physically driving an autonomous vehicle for billions of event free kilometers or covering every available collision scenario imaginable through a theoretical testing plan there could always be scenarios which are unforeseen.

The use of real world collision data in test design is well founded. There are even examples where this approach has led to a huge increase in consumer knowledge and trust. The EuroNCAP testing program is one such example where evidence-based tests have been implemented with the outcome being safer cars, greater public awareness and crucially trust in the system [6].

When searching for evidence in any dataset of real world collisions it is important to understand what you are looking for and how this might inform your research. There are almost no situations where the data will be explored in an ad hoc manner or browsed through in an unstructured way. Most uses of in-depth collision data will begin with a research question. This question forms the focus of the search and defines the boundaries of what information is required. For example; the research question “what are the main causes behind single vehicle crashes involving 17-24 year old car drivers?” gives a range of boundaries which contain only the
Autonomous Vehicle Validation – are we just guessing or can we predict the future?

relevant collisions within the wider population of all collision types. The four degrees of freedom in this case are: collisions with; (i) collision causes available, (ii) single vehicle crashes, (iii) road users aged between 17 and 24yrs and involving (iv) cars. Figure 1 illustrates graphically the population of available collisions within a typical in-depth dataset, in this example the collisions are defined by a ten by ten square with an area representing 100% of all available collisions.

![Figure 1 – Population of available collisions](image1.png)

Overlaying the figure with only the relevant collisions for the research question it is possible to identify the population of collisions that fit the hypothetical research question. Figure 2 shows a representation of the final collision selection from the total sample available. In this hypothetical example the total percentage of cases remaining after the selection process would be 2%; using actual in-depth collision figures drawn from the DfT RAIDS database [7] this would represent a collision sample of 336 from a total sample of 16785.

![Figure 2 – Final selection for RQ](image2.png)

In the example of autonomous pod vehicles this reduction in unrelated crash data is more difficult to achieve. Without knowing anything about the vehicle capabilities, the environment it will exist in, the users who will interact with it or the type of service it will provide it is extremely difficult to define a research question and
Autonomous Vehicle Validation – are we just guessing or can we predict the future?

consequently to identify a targeted population of suitable collision examples.

The method planned through the Capri project is to accurately define the degrees of freedom under which the search will be conducted; the result of any subsequent search will still ultimately not contain any autonomous pods however the data should look like autonomous pod collisions. Within the Capri project it is possible to see what an autonomous pod is like and what it is intended to do; from this the pod can be defined over a wide range of descriptors which can be found in the collision data.

Descriptors that define autonomous pods may be found in terms of their capabilities (acceleration, speed profiles etc.) the environments they use (road classifications, mix of traffic types etc.) and type of expected collision (crash type, impact speeds etc.). The outcomes of each descriptive element can be illustrated on a radar graph, of which each axis provides the values and the area contained within the graph providing an overall ‘envelope’ that describes the road user type. Figure 3 shows a simplified radar graph encompassing the descriptors of (i) acceleration, (ii) road type, (iii) mean ∆v (collision speed), (iv) braking, (v) traffic mix, (vi) speed limit range. The area shaded in the following example shows the typical ‘envelope’ for a passenger car.

Figure 3 – Radar plot showing traditional passenger car envelope

Adding in further road user types in the same manner provides a clear comparison of the different amounts of data that could be available within an in-depth collision dataset. The following image (figure 4) shows the same envelope for passenger cars but with the addition of separate envelopes for cyclists and mobility scooters.
Autonomous Vehicle Validation – are we just guessing or can we predict the future?

Figure 4 – Radar plot showing traditional passenger car, cyclist and mobility scooter envelope

This technique also shows the crossover between the different road user types and allows data from all road user groups to inform the research question, this is especially important when looking for evidence to support vehicles which are not present in the in-depth collision data such as autonomous pods. In other words, this technique does not exclude collisions that do not explicitly fit the profile (i.e. not a pod vehicle) but instead includes all road users and all collision types that fit the general envelope (i.e. looks like a pod vehicle). It is therefore possible to provide a potential envelope for autonomous pods; Figure 5 graphically illustrates the expected population of in-depth collisions that reflect autonomous pods despite no such vehicle being present within the dataset.

Figure 5 – Radar plot graphically illustrating the expected population of in-depth collisions that reflect autonomous pods despite no such vehicle being present within the dataset.
2.1 Example Outcome

As an example of how the sampling methodology selects relevant cases the following collision from a UK dataset demonstrates how a non-autonomous standard vehicle collision can reflect an autonomous pod collision:

*V1 MPV attempts to turn right out of private road onto major public road with shared tram facilities. V2 tram approaches from V1 right at ~20mph. View of V2 approach obscured by parked van directly adjacent to right of V1. V1 edges slowly out onto major road (<1mph). V2 driver fails to see front of V1 emerging from behind parked van and N/S of tram contacts N/S front of V1. Damage only, non-injury collision.*

This example demonstrates that the vehicle paths and characteristics reflect well the type of behaviour and use an autonomous pod will need to contend with. The example collision provides a clear framework from which to base both physical and simulation testing scenarios. In this case either (or both) vehicle(s) can be assumed to be the pod in scenario generation.

3 DISCUSSION

The work outlined within this paper describes the very first steps in defining and determining a relevant collision dataset for autonomous pod testing scenarios. There is currently a need for a robust and detailed dataset for this purpose, analogous to the way in-depth collision datasets have informed the development of regulatory tests for human driven vehicles. There is little doubt that autonomous vehicle development and testing will continue, however it can be seen that specific test scenario generation has until now been based on
a more theoretical approach or from the outcomes of real world testing. These methods, although valid in the complete development lifecycle, could miss critical information and could ultimately damage the perception autonomous pods have on consumers.

The approach outlined has some key benefits over the current testing methods. These include:

- Providing a clearer dataset for further analysis into autonomous pod scenario setting.
- More validity in scenario setting – scenarios are underpinned by data on collisions that actually happened rather than a purely theoretical approach.
- Closer alignment with the real-world pod capabilities – for example it will not be necessary to undertake testing scenarios that are beyond the scope of the pod use.

By employing methods to extract all relevant information from the datasets already at our disposal it will be possible to fill in existing gaps and develop more robust testing scenarios. This step will, in turn, also increase consumer confidence in the testing and safety outcomes of these vehicles as the real-world basis of the testing scenarios are more visible and understandable.

In the initial stages the proposed methodology for identifying pod testing scenarios will need to be set with large confidence margins on the descriptor values. This will ensure that a wide enough net is cast to capture all relevant collisions in the real-world datasets. It is likely that further refinement on each descriptor will be required to tune the model as this approach represents the ‘first best guess’ of an iterative process.

REFERENCES


[2] https://gateway-project.org.uk/


Understanding Trust in an AV-context: A Mixed Method Approach

Fredrick Ekman, Chalmers University of Technology, Sweden, ekmanfr@chalmers.se, Mikael Johansson, Chalmers University of Technology, Sweden, I.C. MariAnne Karlsson, Chalmers University of Technology, Sweden

ABSTRACT

Trust is a fundamental part of technology acceptance as well as an important factor for creating a positive user experience with Automated Vehicles (AVs). In order to fully understand users' trust in AVs it is important to consider the cognitive processes by which humans develop trust. We argue that a deeper understanding of these processes can be elicited by using a convergent mixed method design. The method design described in this paper was created during an experimental study investigating the effect of AV's driving behaviour on users' trust. The design consists of five data collection methods, three qualitative and two quantitative, used to collect data during and after test runs with an AV. The results show that the different methods elicited responses that may indicate different cognitive processes. The methods used during the test runs produced more affective and analogical responses while the methods used directly after each of the test runs generated more analytic responses. The last method, introduced after the completion of all test runs, produced a more mixed result. The participants elaborated on their earlier responses and sometimes turned their affective responses into analytic or analogical explanations. Hence, by combining and utilizing the strength of different data collection methods, more rich data was elicited on the trust formation process and thereby creating a more nuanced picture of users' trust in automated vehicles.

Keywords: Trust, Human-Machine Interaction, Automated Driving, Automated Vehicles, Mixed-Method Design.

1 BACKGROUND

As technology progresses within the area of semi- and fully automated vehicles, it has become increasingly important to consider and understand trust. It is one of the fundamental factors to create acceptance (Ghazizadeh et al., 2012) as well as a cornerstone to create a positive user experience (Waytz et al., 2014). However, trust is a complex concept making it difficult to assess and available methods are few, especially in the area of automated vehicles (AVs). In order to fully understand users' trust in AVs it is important to understand the cognitive processes that govern trust. It has been suggested that trust involves three different cognitive processes (Lee & See, 2004): (i) the affective process which is the emotional process, i.e. feeling trust; (ii) the analogical process which is connecting earlier and familiar experiences and using them to assess the trustworthiness of an agent (in this case the AV) based on differences and similarities, and finally, (iii) the analytic process which is to rationalize around the agent’s trustworthiness. The affective process is the most influential and fundamental trust process for user behaviour, affecting both the analogical and analytic processes of trust. It is also the least cognitively demanding process. The analogical process is used when information about an agent is lacking and earlier experiences (with similar agents) are used to assess the trustworthiness whereas, the analytic process is the logical argumentation of an agent’s trustworthiness in which the information about an agent is evaluated (Lee & See, 2004).
With this paper, we argue for a better understanding of drivers’/users’ trust in AVs through using a convergent mixed method design (cf. Creswell & Plano Clark, 2017), capturing a more nuanced picture with the aim of exploring the respective cognitive processes of trust in an AV-context.

2 METHOD

The design of the method evaluated was created in connection to an experiment investigating how the behaviour of an AV (in terms of acceleration, deceleration, lane positioning etc.) affects the user’s trust. The experiment was conducted on a test course where 18 participants experienced two consecutive AV-test runs with two different driving behaviours. The vehicle had a Wizard of Oz-setup, i.e. it was controlled by a (to the participant) hidden test driver. The test course included a rural- and as well an urban section. Each test run took approx. 15 minutes to complete and included seven ‘critical’ traffic situations. The seven situations were stopping for a red light, overtaking a moving vehicle, stopping for a person waiting to cross a zebra-crossing, passing a cyclist, driving onto a highway, passing oncoming traffic and driving through a roundabout.

2.1 PARTICIPANTS

The 18 participants taking part in the experiment were 10 men and 8 women between 20 and 55 years old (mean 36.7; SD=11.1). The participants were recruited from the Gothenburg city area (on the west coast of Sweden) through a newspaper advertisement. The only inclusion criterion was a valid driver’s license.

2.2 DRIVING BEHAVIOUR

The two distinctly different driving behaviours consisted of a ‘defensive’ behaviour and an ‘aggressive’ behaviour. The respective AV-driving behaviours differed in several ways, for example in acceleration, deceleration and distance kept to other objects (Table 1). All participants experienced both behaviours but in different orders, i.e. half of the participants started with the ‘defensive’ and half started with the ‘aggressive’ driving behaviour.

<table>
<thead>
<tr>
<th></th>
<th>Defensive Driving Behaviour</th>
<th>Aggressive Driving Behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starting &amp; stopping</td>
<td>Keep the vehicle rolling (avoid standstill)</td>
<td>Start &amp; stop (comes to full stop)</td>
</tr>
<tr>
<td>Acc./Decell. Pattern</td>
<td>Avoid heavy acc./decell.</td>
<td>Heavy acc./decell.</td>
</tr>
<tr>
<td>Lane positioning</td>
<td>Early indicate right or left turn (through positioning in lane)</td>
<td>Indicate late right or left turn (through positioning in lane)</td>
</tr>
<tr>
<td>Distance to objects</td>
<td>Keep longer distance (lateral &amp; longitudinal) to other objects</td>
<td>Keep shorter distance (lateral &amp; longitudinal) to other objects</td>
</tr>
</tbody>
</table>

Table 1 – AV driving behaviours.
2.3 PROCEDURE AND DATA COLLECTION

Data collection was performed during two different phases; a peri-trial phase and a post-trial phase (see Figure 1).

The peri-trial phase included two parts and four data collection methods. The first part was conducted during the test run in the AV, and included a Likert scale on trust combined with a Think-aloud procedure (cf. Charters, 2003). The participants rated what level of trust they felt in each of the seven critical situations (i.e. passing a cyclist, stopping for a red light, stopping for a pedestrian crossing the street etc.) and then elaborated on why they felt this way. The second part was completed directly after each test run using a Trust Questionnaire (based on Jian et al., 2000) and a method referred to as the ‘Trust Curve’, an adaptation of the UX-curve (Kujala et al., 2011). The Trust Questionnaire focused on assessing the participant’s overall trust in the AV during each of the respective test runs. The participants were then asked to draw a curve symbolising their level of trust during the test run and to mark out situations that they experienced affected their trust in the AV the most. The post-trial phase, finally, included an interview with the participant, where the ‘Trust Curve’ was used as a mediating tool to allow the participant to further reflect on and discuss the levels of trust in the AV during specific situations as well as their overall trust.

3 RESULT

Initial analyses show that during the first part of the peri-trial phase, the Likert Scale and the Think-aloud procedure provided more affective and analogical responses from the participants regarding the trustworthiness of the AVs. Some of the affective responses during the test described a general feeling of trust during a situation, such as: “It felt good, was calm and comfortable”, while other responses were more detailed, for example: “The trust was high. I am starting feeling comfortable now, I didn’t even care to look right because it felt like nothing was going to happen”. Sometimes the affective responses were of a more dual character, as when one participant expressed a high level of trust during a situation, even though the vehicle had clearly committed what was experienced as a traffic violation, stating that: “It didn’t stop at the stop sign, which I feel is strange but I still believes that it understands”. Another participant elaborated more on the affective response by claiming that: “On a logical level, I would like to say that I had the same trust, but the experience still felt less confident”.

Figure 1 – Data collection methods during peri-trial phase and post-trial phase.

The peri-trial phase included two parts and four data collection methods. The first part was conducted during the test run in the AV, and included a Likert scale on trust combined with a Think-aloud procedure (cf. Charters, 2003). The participants rated what level of trust they felt in each of the seven critical situations (i.e. passing a cyclist, stopping for a red light, stopping for a pedestrian crossing the street etc.) and then elaborated on why they felt this way. The second part was completed directly after each test run using a Trust Questionnaire (based on Jian et al., 2000) and a method referred to as the ‘Trust Curve’, an adaptation of the UX-curve (Kujala et al., 2011). The Trust Questionnaire focused on assessing the participant’s overall trust in the AV during each of the respective test runs. The participants were then asked to draw a curve symbolising their level of trust during the test run and to mark out situations that they experienced affected their trust in the AV the most. The post-trial phase, finally, included an interview with the participant, where the ‘Trust Curve’ was used as a mediating tool to allow the participant to further reflect on and discuss the levels of trust in the AV during specific situations as well as their overall trust.
In analogical responses the participants assessed the trustworthiness of the AV through comparing familiar driving styles with the driving styles experienced during the experiment. Some participants compared one of them to the driving style of people they knew. One participant explained: “It was more aggressive, it felt like that mate who thinks it’s fun to drive fast and swing back and forth in the lane” and another said: “I have a son at home who is taking his license now. It was more like his driving style, maybe not the risk-taking but the jerkiness and the lane positioning”. Yet another participant compared the driving style to the driving style of a professional driver: “This time I think it drives like an experienced taxi driver that drives comfortably. It was a bit more stressed the first time, I prefer the calmer one”. However, some analogical responses were of a more abstract character, comparing the driving behaviour with other familiar experiences, for example to a rollercoaster ride: “It was fun, like the ‘grandpa cars’-ride at Liseberg” (a rollercoaster ride located in an amusement park in Gothenburg).

During the second part of the peri-trial phase, the Trust Questionnaire resulted in a more general view on the participants’ trust for the AV whilst the ‘Trust Curve’ produced more analytic responses regarding the AVs’ trustworthiness. One participant described the driving behaviour’s effect on his/her understanding of the system’s intentions as: “There was a distinct retardation (braking), which indicated that the system detected that there was something there, and then it braked”. Another participant explained how this understanding of the system’s intentions increased his/her perceived trustworthiness of the system: “When the pedestrian crossed the street, I got a feeling that the car had actually seen the pedestrian and then braked softly and stopped. That is this trustworthiness that I think was lacking before”.

Finally, the post-trial phase combination of ‘Trust Curve’, used as a mediating tool, and in-depth interviews allowed the participants to reflect on their responses. The nature of the responses during the post-trial phase were mixed in that the participants sometimes described affective responses referring to the ‘Trust Curve’ and turning them into analogical and analytic explanations.

4 DISCUSSION AND IMPLICATIONS

A number of studies have used different types of questionnaires to measure trust in technology, in general and AV in particular (e.g., Merritt et al., 2013; Haeuslschmid et al., 2017). Some, but in comparison few, have used (also) for example interviews to gather more qualitative data (e.g., Xu et al., 2014). However we argue a more comprehensive approach, including a mixed method design, to fully understand the processes of trust in an AV context.

Affective, analogical as well as analytic responses were elicited in the trial. However, different data collection methods extracted (to a certain degree) responses of different character. The methods used during the test runs extracted the majority of the affective and analogical responses, while the methods used directly after the respective test runs produced mostly analytic responses. The method used after both test runs, i.e. during the post-trial phase, elicited mixed responses and sometimes made the participants further elaborate on their earlier answers, and turning affective responses into analogical or analytic explanations.

The first data collection methods, i.e. the combination of a Likert scale and think-aloud procedure, were chosen in order to obtain direct and unfiltered answers from the participants. One explanation to the majority of affective
answers being collected during the first combination of methods (part 1) is probably that this data was collected ‘in situ’, i.e. during and directly after a ‘critical’ traffic situation. The Likert scale provided a trust measurement, indicating the participant’s level of trust at the time, while the think aloud procedure allowed the participants to express their spontaneous reactions and opened up for immediate emotional responses, triggered by the situations. These affective responses would probably have been fewer if data collection had taken place after the test runs only, when the immediate emotional response may have been forgotten – affect fades faster than cognition (Norman, 2009) or is combined into an overall experience, something argued by the ‘peak-and-end’ rule (e.g., Fredrickson, 2000) which states that the most affectively intense event (e.g., a particular situation during a test run), together with the affect experienced at the end of a sequence (e.g., at the end of test run), affect the entirety of an experience (e.g., the entire test run). Hence, a questionnaire used at the very end of the test run may not capture the nuances of affective responses associated with different situations making it more difficult to understand participants’ trust.

The Trust Curve used directly after each test run (in part 2 of the peri-trial phase) produced mostly analytic responses. By letting the participants draw a curve from memory of their perceived level of trust and also write down the episodes that affected their trust the most, the most prominent situations were annotated. At the same time as ‘peak-and-end’ rule to some extent contradicts the idea of retrospective information retrieval, Karpanos et al. (2010) as well as Kujala et al. (2011) argue that such recalled memories are important since they are the most meaningful to the users. The sketching of the trust curve itself may also be of importance in recalling memories, since earlier studies have shown that when sketching chronological curves participants recalled more experiences than participants using no form of sketching (Karpanos et al., 2012). The number of analytic responses could be explained by the fact that the method allowed the participant to also reflect on the test run as a whole as well as how the individual situations shaped the overall experience. Here, the act of actually drawing the Trust Curve may also be an important characteristic of the method as such generative elements are argued as essential in order to elicit deeper knowledge about experiences (cf. Sleeswijk Visser et al., 2005).

Finally, even though the different data collection methods used in this study extracted responses of different character, it is not possible to say which particular method should be used to elicit a certain type of response. Rather the respective methods seem to complement each other in a way that enables the elicitation of richer data. This data could in turn be analysed to further explore which cognitive processes are initiated during different parts of the users’ interaction with AVs (such as during the different situations) as well as in what way.

The implications are that to develop a deeper understanding, it is not sufficient to use one, single, data collection method but instead a combination of methods to elicit different aspects of the users’ trust, both immediate as well as more reflective responses. Further implications are that data should not only be collected at the very end of a trial only but be complemented with data collection also during a trial, in particular in relation to events that may influence and contribute to a user’s overall experience. The methods used in the study differ as to their respective affordances and constraints, some of which appropriate for immediate response (i.e. non-obstructive), others sensitising and supporting reflections. Hence, utilizing a convergent mixed method design approach could facilitate a more nuanced image of users’ trust for AVs, by exploring the different cognitive processes governing trust in different situations.
REFERENCES

Charters, E. (2003). The Use of Think-aloud Methods in Qualitative Research; An Introduction to Think-


Fredrickson, B. L. (2000). Extracting Meaning from Past Affective Experiences: The Importance of Peaks,


Experiences Over Time: Memory of Actuality? CHI ’10 Extended Abstracts on Human Factors in Computing
Systems, 4075-4080.

International Journal of Human-Computer Studies, 70(11), 849-865.


50-80.

Norman, D. A. (2009). THE WAY I SEE IT: Memory is More Important than Actuality. Interactions, 16(2), 24-
26.


Xu, J., Le, K., Deitermann, A. & Montague, E. (2014). How different types of users develop trust in
technology: A qualitative analysis of the antecedents of active and passive user trust in a shared technology.
Applied Ergonomics, 45, 1495-1503.
Vehicle Movement and its Potential as Implicit Communication Signal for Pedestrians and Automated Vehicles

Claudia Ackermann, TU Chemnitz, Germany, claudia.ackermann@psychologie.tu-chemnitz.de, Matthias Beggiato, TU Chemnitz, Germany, Luka-Franziska Bluhm, TU Chemnitz, Germany, Josef Krems, TU Chemnitz, Germany

ABSTRACT
An important challenge for automated vehicles will be the coordination of their own actions with the behaviour of other road users such as pedestrians. A possible solution for pedestrian-automated vehicle interaction could be explicit interfaces that convey additional visual or auditory signals (e.g., by projection). However, it is argued that light and weather conditions could significantly affect the functionality of such interfaces. Implicit communication signals such as the movements of a vehicle are already used by pedestrians, so that they are able to cross the street without any further communication. Thus, the study addresses the question of which parameters influence the recognition and classification of vehicle movements into acceleration, deceleration and constant speed driving. Therefore, we implemented an experimental video simulation in which the independent variables vehicle speed (20 and 40 km/h), daylight (morning, dusk, evening), onset of the movement change (early, late) and acceleration rate (positive, negative, none) were varied. The task of the participants (n = 33) was to indicate by pressing a button when they had detected changes in the approaching car movement. Further, they had to decide what kind of movement it was (deceleration/acceleration). In the results we expect answers to which parameters or parameter combinations have a positive or negative influence on the recognition of deceleration, acceleration or constant speed driving. First exploratory analyses reveal an influence of speed and deceleration rate on detection time. The influence of daytime and the onset of deceleration seems to be rather subordinate or not clear yet.

Keywords: implicit communication, communication pedestrian and automated car, recognition of vehicle movement.

1 INTRODUCTION
The transport systems of the future will be determined by an ever-increasing proportion of automated vehicles (AV). Current market predictions forecast 21 Million AV sales by 2035 (IHS Automotive, 2016). In addition to this promising outlook on the future of automated driving, there are still many challenges, like for example the regulation of autonomous behaviours in terms of legal issues (Gasser, 2016). A further challenge which AVs has to face is how to communicate with other road users, like pedestrians, where there is still comparatively little research (Pillai, 2017; Rasouli, Kotseruba & Tsotos, 2017). One approach to configure communication between pedestrians and AVs is to incorporate external interfaces, often in terms of visual communication via LED displays or projections (see for example Lagström & Malmsten Lundgren, 2015; Nilsson, Thill & Ziemke, 2015). In complex, busy traffic environment with sometimes bad weather or light conditions, it is questionable whether the comprehensibility of such external interfaces is ensured in any case (Pillai, 2017; Rasouli et al., 2017; Risto, Emmenegger, Vinkhuyzen, Cefkin & Hollan, 2017). In current traffic, the assessment of vehicle movement seems to be an implicit and functional strategy for cooperative communication among vehicles and
pedestrians (Müller, Risto & Emmenegger, 2016; Pillai, 2017; Rasouli et al., 2017). Implicit communication is familiar to the involved interaction partners and also visible from different perspectives. This allows pedestrians to cross a road without having a direct communication with the driver, such as at night (Rothenbücher, Li, Sirkin, Mok & Wu, 2016). The observational findings of Rothenbücher et al. (2016) and Risto et al. (2017) suggest that different types of vehicle decelerations or accelerations in front of the pedestrian lead to behavioural adjustments, such as sidestepping of the pedestrian. Pillai (2017) conducted a virtual reality experiment to investigate several types of deceleration and pedestrian’s feeling of interaction comfort. Most comfortable deceleration types where characterized by smooth speed reduction that starts rather further away from the pedestrian. Deceleration types with a mixture of deceleration and acceleration were assessed to be most uncomfortable. One question imposed at this point is, on which parameters do people determine vehicle movement at all, before a behaviour change or a comfort evaluation can take place? Beggiato, Witzlack, Springer and Krems (2017) published a video simulation study, where they investigated the effect of daytime and speed of the approaching car on participant’s assessment of the latest moment for crossing the street comfortably. Indirectly it could be concluded, at which time the participants would expect a brake initiation of the vehicle. Results indicate, there is no “one-fits-all” solution for comfortable braking, but this study revealed the two important parameters vehicle speed and daytime. A study which specifically dealt with the identification of vehicle braking, examined the deceleration rate (5 m/s² vs. 3.5 m/s²) in addition to different vehicle speeds (50 km/h vs. 30 km/h) to have an impact on detection time (Petzoldt, Schleinitz & Banse, 2018). The authors found that participants had shorter detection time for lower speed and higher deceleration rate. The present study therefore focuses on the detection of vehicle movement. New here is the investigation of different parameters and their combinations in a common experimental setup. The aim was to investigate which parameters an parameter settings have an influence on the detection performance. In addition, the nature of the vehicle movement was investigated by including both deceleration, acceleration and constant driving in the experiment. We chose to include vehicle acceleration and constant driving to investigate every possible way of vehicle behaviour concerning speed. Further, we also addressed acceleration to explore, whether these type of movement could be useful as implicit communication signal.

2 METHOD
2.1 Participants

A total of 33 persons participated in the study. The sample consisted of 22 women and 11 men ranging from 18 to 47 years (M = 23 years, SD = 5.38) who were all students at Chemnitz University of Technology. Slightly less than half of the subjects had normal vision (45 %), the remainder used glasses or contact lenses to correct vision (55 %). Participants managed their daily routes, primarily on foot (76 %) or by public transport (45 %), especially in case of shorter distances between one and 20 kilometres (79 %).

2.2 Study Design

This experimental study was based on a video-simulation using a within-subject design. We varied the independent variables speed (20 km/h vs. 40 km/h) of the approaching vehicle. Furthermore, the independent variable acceleration had seven degrees which had positive, negative and neutral values (+/- 5, 3.4, 1.5 and 0
m/s²; based and extended from Petzoldt et al., 2018). We could simulate three types of motion by that, namely acceleration, deceleration and constant driving. The third independent variable was the onset of adjusted vehicle movement. Here, we implemented an early onset (between 3.5 and 4.5 s before reaching the target position) and a late onset (between 2 and 3 s before reaching the target position; both time gaps were based on findings by Beggio et al., 2017). We preferred these onset time gaps over fixed points of onset, because we were able to reduce the subject’s probability to guess the right answer. The target position was a yellow line on the street surface, which we used to calculate distances for the acceleration onset and profiling (Figure 1). This line was not visible to the participants during the simulation. As fourth independent variable, we used the daylight which was varied between in the morning, dusk and evening light condition. We measured the reaction time between onset of the acceleration or deceleration and detecting the movement of the vehicle as dependent measure. Furthermore, we recorded the answer of the participants which kind of movement they recognized. In total, there were 2x7x2x3 = 84 trials which were shown in a randomized order.

**Figure 1 – Recording perspective and yellow line**

### 2.3 Video Material and Simulation Software

The experiment included prerecorded real-world videos on a public parking area at Chemnitz University of Technology. All videos were recorded at exactly the same place and one day, between 11:13 AM and 19:25 PM. These two time stamps represent the daylight conditions in the morning and dusk. The daylight condition in the evening was generated by darkening the video at 19:25 PM using the software VirtualDub. The recording was done using a GoPro Hero 4 camera with Full HD resolution of 1920×1080 pixels and 120 frames per second. The camera was placed at a height of 1.60 m and 50 cm left of the roadside (Figure 1). This position was assumed to be similar to the perspective of a pedestrian intending to cross the street. The oncoming vehicle was a small passenger car (white smart electric drive), driven by an investigator at a constant speed of 20 km/h. To ensure precise vehicle speed in the videos, markers like the yellow line, were placed at the street surface. For the video recording session it was made sure that no other moving objects such as other vehicles or pedestrians passed the scene. The scenes were shot in calm weather, so no movement of the trees and plants in the video can cause irritation. Using the markers and the video editing software LabVIEW 2015, all videos were accelerated or decelerated to get an exact and constant speed, acceleration rate and acceleration onset depending on the experimental condition. Due to these precautionary measures, vehicle speed, acceleration rate and onset of accelerations could exactly be manipulated without artificial side-effects in the simulation software by regulating the playback speed of the videos. Using LabVIEW 2015, the videos, instructions and messages were presented and the detection time and responses of the subjects were recorded.
2.4 Procedure

Before the start of the simulation, all participants were informed about the scope and procedure of the study and anonymity was guaranteed. Thereafter, socio-demographic and pedestrian behaviour information were collected (based on Papadimitriou, Yannis & Gilou, 2009). The instruction was to press a defined key, when the participant noticed the kind of vehicle movement. They were instructed to press the key, when they were as sure as possible in their decision about the vehicle movement. Further, they were instructed not to press the key, when the vehicle is driving with constant speed. After pressing the key or watching the video to the end, they were asked to decide what kind of vehicle movement they recognized (Figure 2). After the instruction, two test trials were presented. Subsequently, the 84 experimental trials were presented in randomized order. After each trial, a dialog message with an OK Button was shown to give participants control over the start of the next trial whenever they were ready.

![Participant's view on video simulation and answer screen](image)

Figure 2 – Participant’s view on video simulation and answer screen

3 RESULTS

As part of the data preparation, a subject was removed due to instructional adverse behavior. Only those data were included in the calculations for which both the vehicle movement was correctly detected and whose detection time was not prior to the initiation of the respective change in movement. Since there is no reaction time for the condition with constant speed, this condition was not included in the analyses.

3.1 Detection of Acceleration and Deceleration

In a first step, we analysed the proportion of false responses for both types of vehicle movement. A high amount of participants gave wrong answers in the acceleration condition (on average per trial, only 13 participants gave correct answers) in comparison to the deceleration condition (on average per trial, 27 participants gave correct answers). For that reason, all data regarding the acceleration condition were excluded from further analyses. But, a closer look at the deceleration condition also revealed problematic trials. Those trials were characterized by higher speed (40 km/h), a low deceleration rate (1.5 m/s²) and a late deceleration onset time (between 2 and 3 s). In that trial, only about 13 percent of participants gave the correct answer, independent from daytime. For that reason, we decided to describe our results in an explorative way first. Within the first steps of analyse, we focused on general tendencies among our parameter. Thus, we found almost no impact of daytime, which is presented as first result, followed by vehicle speed, onset of deceleration and deceleration rate.

3.2 Daytime
We averaged here the trials for each daytime condition separated by speed (Table 1) across all deceleration trials. When comparing the mean detection time, it becomes obvious that daytime seems to have probably not a big impact, but rather speed.

Table 1 – Mean detection times separated by daytime and speed (standard deviation enclosed in brackets)

<table>
<thead>
<tr>
<th></th>
<th>Morning</th>
<th>Dusk</th>
<th>Evening</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 km/h</td>
<td>1.59 s (0.43)</td>
<td>1.72 s (0.64)</td>
<td>1.63 s (0.57)</td>
</tr>
<tr>
<td>40 km/h</td>
<td>2.72 s (0.71)</td>
<td>2.65 s (0.86)</td>
<td>2.51 s (0.52)</td>
</tr>
</tbody>
</table>

3.3 Vehicle Speed

To analyse this parameter, we averaged across all deceleration trials representing the lower (20 km/h) and higher (40 km/h) speed condition. We found shorter reaction times for any trials with lower speed (Mean = 1.61 s, SD = 0.66) than for higher speed (Mean = 2.60 s, SD = 0.86). So, it seems as if participants were able to detect deceleration manoeuvres faster, when the car approached with lower speed.

3.4 Onset of Deceleration and Deceleration Rate

Here, we averaged the trials with late and early onset of deceleration with no regard to daytime, but separated by deceleration rate. Because of missing values for the higher speed condition, we only describe results referring to the lower speed condition (20 km/h). Table 2 lists the mean reaction times for deceleration rate in relation to onset of deceleration. Tendencies can be assumed for the impact of deceleration rate; the stronger the deceleration the shorter detection times can be found. The influence of the onset of deceleration seems to be not that stringent. Participants showed shorter reaction times when the vehicle started to decelerate further away from them (3.5-4.5 s), except for the medium deceleration rate.

Table 2 – Mean detection times separated by deceleration rate and onset of deceleration (standard deviation enclosed in brackets)

<table>
<thead>
<tr>
<th>Deceleration Rate</th>
<th>-1.5 m/s²</th>
<th>-3.4 m/s²</th>
<th>-5 m/s²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Onset of Deceleration</td>
<td>2-3 s</td>
<td>2.60 s (0.96)</td>
<td>1.37 s (0.45)</td>
</tr>
<tr>
<td>Deceleration</td>
<td>3.5-4.5 s</td>
<td>2.14 s (0.59)</td>
<td>1.39 s (0.41)</td>
</tr>
</tbody>
</table>

4 DISCUSSION

The aim of the study was to identify parameters that describe the detection of vehicle movement and thus can be used as an implicit communication tool for pedestrians and automated vehicles. On the basis of the previous descriptive exploration of the data, it can be assumed that acceleration processes are recognized much worse than braking. The lack of acoustic cues (engine noise), which are primarily noticeable when accelerating, might be an explanation here. Alternatively, a method effect could be conceivable, because in the accelerating conditions the absolute duration of the videos was shortened by speeding up the video, which could have been critical in particular for the late onset of acceleration. Further, the difference between the two onset ranges is only 0.5 sec. Thus it might be possible there was not enough contrast between the onset ranges to be perceived
as distinct levels. We found tendencies, which confirm the results of Petzoldt et al. (2017), who also found shorter detection times for higher deceleration rates and lower speed. We could not assume daytime to have an unambiguous influence on detection time, contrary to the effects of Beggiato et al. (2017b). Onset of deceleration seems to be an influencing factor on detection time, but probably interacts with other aspects, like deceleration rate. It should be noted that deceleration is the more appropriate implicit communication solution compared to accelerations. This might be confounded by the two-stage response process we conducted. In further experiments, it could be helpful to use two buttons (one for detection of acceleration, one for detection of deceleration), then reaction time and choice can be measured simultaneously. Further analyses must address the characteristics of the individual parameters in order to derive substantiate information regarding algorithm development in AVs.

ACKNOWLEDGEMENTS

This project was funded within the Priority Program 1835 “Cooperative Interacting Cars” of the German Science Foundation DFG.

REFERENCES


DRIVER BEHAVIOUR AND TRAINING
Proceedings of the 6th Humanist Conference, The Hague, Netherlands, 13-14 June 2018

Education and training of future car drivers in Flanders

Roel De Klerk, VSV, Belgium, roel.deklerk@vsv.be, Kim Jacobs, VSV, Belgium.

ABSTRACT

Despite a positive evolution over the last decade, car drivers aged 18 to 25 remain over-represented in road accident statistics in Flanders. Therefore, the Flemish road safety policy includes specific objectives to lower the number of fatalities in young, novice car drivers and focuses on improving the driver training. Commissioned by the Flemish minister for Transport and together with other partners in the field, the VSV (the Flemish Foundation For Traffic Knowledge) has elaborated an extensive revision of the driver training in Flanders. This revision has been adopted by the Flemish government in 2016 and is currently in the course of implementation.

Keywords: Road safety, novice drivers, driver training.

1 ROAD SAFETY IN FLANDERS

1.1 General

Based on the most recent preliminary data (2017), Flanders counts 48 traffic fatalities per million inhabitants, which is almost the same as the EU average (49). The number of road deaths for 2017 is estimated at 290, and there was a record low for all accident indicators. Although the death toll has considerably decreased over the last decade, further measures are needed to reach the goal of maximum 200 fatalities in 2020, adopted by the Flemish government (Figure 1).

1.2 Car occupants

Despite recent record lows in fatality numbers, car occupants remains the biggest group in the accident statistics (Figure 2). Especially young car drivers aged 18 to 25 are over-represented, although the general positive evolution also applies to them (Figure 3). The Flemish road safety policy therefore includes specific objectives to lower the number of fatalities in young car drivers, and focuses on improving the driver training,
Education and training of future car drivers in Flanders among other measures (Vlaamse overheid, 2016).

Figure 2 – Number of on-the-spot fatalities according to road user type, Flanders

![Figure 2](image)

Figure 3 – Evolution of killed and seriously injured young car drivers, Flanders, 2005-2016

![Figure 3](image)

Source: FPS Economy, GD Statistics – Statistics Belgium
2. LIFELONG LEARNING CONCEPT

Traffic education in Flanders already starts in the kindergarten. To enable and promote lifelong learning in traffic, VSV has developed a learning path and steps that ensure that pupils learn in a step-by-step manner how to behave safely and independently in traffic as pedestrians and cyclists. In general, the accident risk increases when children or youngsters start using a new mode of transport independently: there is a clear rise in the number of victims among bicycle riders between the ages of 12 to 14 and among moped riders between the ages of 16 to 18. Likewise, the accident risk increases even more sharply in car drivers aged between 18 and 24 (Lammar, 2016).

Driving a motor vehicle requires taking adequate decisions rapidly and without hesitation. The skills needed for this task have to be learned by practicing a lot. A number of skills can be taken over by vehicle technology, but currently this is mostly limited to driver assistance in case of emergency situations. The risk is mainly in the behaviour of the driver, especially among young people who overestimate their own abilities and underestimate the complexity of traffic, leading to unnecessary and dangerous errors. Despite technological advances, education remains indispensable to inform, convince and train young road users into adopting safe attitudes and behaviours in traffic (SWOV, 2016) automobilisten)

3 NEW DRIVER TRAINING IN FLANDERS

3.1 Framework

With the sixth institutional reform that came into effect in 2014, many competencies were transferred from the Belgian federal level to the regions, including competencies related to road safety policy. The Flemish authorities have opted to use those newly acquired competencies to thoroughly review the driver training in Flanders. The aim is to raise the bar for apprentice drivers by requiring more knowledge and skills. In addition, the system of accompanied driving is maintained and better support is provided for this.

VSV chairs the Driver Training working group within the Flemish Road Safety House, the official body that supervises the Flemish road safety policy. The working group includes representatives from all parties involved in driver training in Flanders. In 2014, the Driver Training working group approved the official curriculum for the B driving licence, which forms the basis for the driver training revision.

In 2014, CIECA analysed the effectiveness of different options for the B licence driver training (CIECA, 2014). The results of the analysis served as a basis for policy recommendations by the Driver Training working group that were eventually approved by the Flemish government. The policy recommendations included:
Education and training of future car drivers in Flanders

- the introduction of a phased driver training;
- an extension of the minimum driver training duration;
- the adaptation of the driver exams;
- supporting the supervisors (experienced drivers accompanying the learner driver);
- mutual reinforcement through the combination of professional training and exercise under supervision of an experienced driver;
- the use of a logbook to support the learning process;
- monitoring the cost;
- maintaining the good elements of the current driver training.

In 2016, the Flemish minister for Transport commissioned VSV to further elaborate the legal and regulatory framework of the driver training revision, leading to Flemish government decrees to modify the theoretical and practical exams and the driver training. In the same period, VSV also elaborated a handbook for practice instructors and a handbook for supervisors. In this way, the system of accompanied driving with supervisors is maintained with professional support.

3.2 The road to the B driving licence in Flanders

3.2.1 From 1 June 2017 onwards

As from 1 June 2017, the assessment of the theory exam has been modified. The exam consists of 50 multiple choice questions, and candidates succeed if they have a score of 41/50 or above. In case of a wrong answer, the candidate loses 1 point, but if the wrong answer implies a serious violation of traffic rules or if it is related to speeding, the candidate loses 5 points.

Also the practical exam has been modified. The exam has been extended with a computerized hazard perception test. During the exam on the road, candidates have to drive independently with the help of a satellite navigation system or by following road signs. Candidates must also perform 2 maneuvers, chosen at random from 6 possible maneuvers.

3.2.2 From 1 October 2017 onwards

As from 1 October 2017, the minimum exercise period before candidates can take the practical exam has been extended from 3 months to 9 months. A smartphone app and a practical manual are available for candidates and supervisors, allowing a step-by-step approach of the exercise period and giving examples for concrete exercises. The app also allows tracking the kilometres driven when exercising. Supervisors are obliged to follow a 3 hour introductory course before they are allowed to accompany a candidate.

The content for the course was developed by VSV with a view to pedagogy and a logical approach to the learning process of a learner driver. The basic principle is that human behaviour is the main factor in traffic
Education and training of future car drivers in Flanders

accidents and that proper behaviour should therefore be discussed from the beginning of the course.

Attention is also paid to the importance of exemplary behaviour that parents pass on to their children. The principles of the Goals for Drivers Education (GDE) matrix are incorporated in the training. Participants learn that a good driver must not only be a skilful driver but must also have the right attitude, and that the social context can lead to risky behaviour in learner drivers.

6 to 9 months after passing the practical exam and obtaining their driving licence, candidates have to follow a 4 hour post-licence training. This involves a group discussion and a number of exercises in a training ground, aimed at increasing the understanding of the own limitations as a driver, and the control of traffic risks.

4 CONCLUSIONS

Despite the decreasing trend in the accident figures in Flanders, extra efforts are needed to reach the 2020 objectives set by the Flemish government. The government wants to deploy specific measures for novice drivers, since they still represent an important risk group. The recent state reform offers opportunities to raise the bar for learner drivers and the obtention of the driving licence. It is expected that, partly due to the combination of lifelong learning and the driver training reform, Flanders will be a step closer to achieving its road safety targets by 2020. Further monitoring by the Flemish Road Safety House (VHV) will be necessary to evaluate and if necessary adjust the measures that have been taken.

5 REFERENCES


SWOV (2016). 18- tot en met 24-jarigen: jonge automobilisten. SWOV-factsheet, mei 2016 , Den Haag
CIECA (2014). Category B driver licencing in CIECA member countries and beyond: overview of the category B licencing process for novice drivers in 43 countries worldwide


Martensen, H (2014) @RISK: Analyse van het risico op ernstige en dodelijke verwondingen in het Verkeer in functie van leeftijd en verplaatsingswijze. Brussel, België: Belgisch Instituut voor de Verkeersveiligheid – Kenniscentrum Verkeersveiligheid
Automated feedback on viewing skills lowers accident involvement

Jorrit Kuipers, Delft University of Technology / Green Dino, Netherlands, jorrit@greendino.nl, Peter Wieringa, Delft University of Technology

Keywords: automated feedback, viewing skills, accident involvement, driving simulator, learning theory

Abstract

The risk of being involved in an accident in the first year after licensing is greater for novice drivers who passed their driving exam the first time than for novice drivers who failed their first driving exam. Enhanced training programmes can shorten the duration of training and can raise the passing rate on the first exam, but can also increase accident involvement after licensing. We propose automated feedback on viewing skills can contribute to safe driving after licensing. An intervention was made in a driving simulator curriculum to study the transfer on the first driver exam and retention of driving skills for safe driving in the first year after licensing. A questionnaire was sent to 22,881 former students. The results of 2,439 subjects where used in this study. The driving skills of a control group were compared to the driving skills of subjects who followed driving lessons with automated feedback on viewing behaviour. Analysis of simulator data and questionnaire data showed significant differences between the two groups. Novice car drivers who followed driving lessons on a simulator with automated feedback on viewing skills needed fewer lessons to pass the driving exam. The self-reported accident involvement of this group was 31% lower than the control group and 32% lower than the average accident involvement in the Netherlands. We suggest using automated feedback on viewing skills in driver training before and after passing driver examination to increase road safety.

1 Introduction

The risk of being involved in an accident in the first year after licensing is greater for novice drivers who pass their driving exam the first time than for novice drivers who fail their first driving exam (Renge 1983, Fortsigh et al. 1997, Wells et al. 2008). Enhanced training programmes, like skid avoidance training, can shorten the duration of training and can raise the passing rate on the first exam, but can also increase accident involvement after licensing. The retention of skills necessary for safe driving, showed during the driver exam, is relative low in the first months after licensing. After licensing, the feedback of the driving instructor stops immediately. The loss of the external feedback directly leads to erosion of skills EOS (Kuipers 2014). Novice drivers in the Netherlands are 4 to 6 times more frequently involved in accidents compared to experienced drivers. This phenomenon is reported worldwide. It seems difficult for formal driver training programmes to achieve positive retention on safe driving after licensing (Brown et al. 1987, Brown 1997, Mayhew et al. 1998, Christie 2001, Elvik & Vaa 2004).

Positive effects on safe driving where noted as result of training of recognition of dangerous traffic situations (Vlakveld 2011). We assumed that extra attention for viewing skills can have a positive effect on the retention of safe driving and can lower accident involvement. We proposed using automated feedback on viewing skills
Automated feedback on viewing skills lowers accident involvement

because it is too demanding – and unsafe – for a driving instructor to give consistent feedback on viewing behaviour. We assumed not only the retention of safe driving skills will benefit from automated feedback on viewing skills, but also the transfer of training during examination. To test our hypotheses, we developed and implemented viewing feedback technology in a driving simulator curriculum. A learning theory was constructed that embraces EOS. We compared the driving skill performance of simulator students who were trained with the automated viewing feedback to students who followed driving lessons without the automated feedback. And we performed analysis between subjects to lower self-selection effects of the driving simulator. An online questionnaire was used to research the transfer and retention of safe driving skills.

2 Mental Transition

To have a better understanding of EOS, we constructed a new learning theory, Mental Transition (MT). MT is an abstract model, based on generally accepted learning theories from the field of neural psychology. We suggested that a better understanding of the biological principles underneath information processing in the human brain could result in improved retention of safe driving. MT has two pillars. The first one is automation of skills; the transition from information processing from the short term memory (STM) to the long term memory (LTM). New information is processed in the STM. STM is slow memory with a very limited capacity and therefor error prone. The shift from STM to LTM is necessary to quickly process large amounts of information without making mistakes. This mental transition is not stable. In case the permanent nerve structure is not stimulated frequently, the nerve connections become weaker and can vanish. EOS appears, the mental transition rolls back.

The second pillar of MT is based on the parameters that influence the mental transition. The frequency of sensory stimuli and the complexity of information are important parameters influencing the performance of information processing in the brain. Parameter management can optimise the process of mental transition and lower EOS. Speed for example, has strong correlation with frequency and complexity of information. Other parameters we distinguish are motivation (internal and external) and intelligence (information processing capacity). We tuned these parameters to achieve an optimal performance of the human brain related to the learning process. For example, 50 metres before a crossroads, a student receives the instruction to release the accelerator pedal and decrease speed. Normally, drivers maintain speed and use the brake to lower speed. Entering the crossroads at higher speed leads to tunnel vision. The mental effort needed to process information increases. The field of view decrease. Special attention was given to weak stimuli, like applying traffic rules. We tried to associate them with stronger stimuli such as vehicle handling. Slowing down not only results in an increased field of view, but it also supports the learning of skills for applying rules.

3 Automated feedback on viewing skills

In 2009, we introduced automatic facial recognition and visual feedback for learner drivers who followed driver training on a driving simulator (Figure 1).
Automated feedback on viewing skills lowers accident involvement

The learning theory Mental transition and the interface design methodology Data Centered Design DCD (Kuipers 2014) were used to construct an automated feedback mechanism. The feedback mechanism contains an adaptive information management system and a user interface for communication with the driver. The information management system contains information on driving procedures, peer group performance, and mental effort. In case viewing behaviour is part of a driving task, the view assessment is activated. 8 visual areas are specified (front, front sides, inside mirror, outside mirrors, and sides). By showing student a red field in the area where he/she should have looked, in combination of a displayed warning, the student receives necessary feedback to solidify his/her knowledge (Figure 2).

Auditive feedback on the driving speed before entering a crossroads is part of the feedback mechanism for viewing skills. The infrastructure was designed to support the feedback frequency threshold of maximum 1 minute. Within one minute, the student enters the next crossroads and rehearses the driving task. The student receives the instruction and warning to release the gas pedal 50 metres before entering the crossroads. We thought this way, the student could associate the frequency of red areas with the vehicle speed. We tried to teach them the relations between speed and viewing behaviour. After the lesson, scores for parameters like speed before crossroads and viewing behaviour are displayed on screen and reported by email to support self-reflection and motivation to rehearse and improve safe driving skills, like viewing. The safety report shows correlations between parameters involved in safe driving (Figure 3). Students and supervisors can use this...
Automated feedback on viewing skills lowers accident involvement

information to learn/teach the consequences of specific driving styles, like high speed before crossroads results in tunnel vision. In the simulator, the student consistently receives the advice to release the accelerator pedal in case of an unsafe speed. After seeing the safety scores, the student gets a better understanding of the negative relation between speed and viewing. Paying attention to the visual and auditive feedback directly results in higher scores on both parameters.

Figure 3 – Safety report: assessment scores on safety parameters. Correlation between safe speed and view behaviour.

4 Effect study

In 2015, we started an effect study among former students who followed driving lessons on driving simulators with automated feedback. An e-mail was sent between November 9th and 13th, 2015 to 22,881 people whose e-mail addresses were in the databases of driving schools with a Green Dino simulator. In this e-mail, the researcher first introduced herself and explained the subject and main goal of the questionnaire. The importance of the recipient filling out the questionnaire was also highlighted. The (former) driving students were then asked to complete the online questionnaire by clicking on the provided link and answering the questions. As a reward, 10 x 2 cinema tickets were promised to be raffled among interested participants. People could also indicate if they would like to be sent a summary of the results once the research had finished. The raffle and sending the summary took place in March 2016.

In December 2015, several driving schools were approached and asked if they could forward the aforementioned e-mail message to their (former) students, so more regular students could be reached and
Automated feedback on viewing skills lowers accident involvement

serve as a control group. To maximise the control group, the researcher also shared the link to the questionnaire on social media. After merging the questionnaire data with simulator data, people who had participated in 0 or 1 simulator lessons were also added to the control group. The online questionnaire was closed on February 10th, 2016.

Eventually, 6,729 people have viewed the questionnaire, 5,142 people started answering questions and 3,761 people actually completed the questionnaire. After inspection of the data, 1,322 completed questionnaires (35.1%) appeared not to be useful for analysis. 2,439 subjects remained, of which 72.2% were simulator students (had lessons on a driving simulator during their driver’s education). Only fully completed questionnaires were taken into account, and only when the participant was in the possession of a driving licence since 2007 or later (because previous research on this topic was about 2007 and before). A short summary of subject data is listed in Table 1.

Participants were removed if they were younger than 18; the legal age for having a driving licence. Because age differed significantly between the simulator group and the control group, age categories were computed to examine separately if necessary. Participants were divided in 3 age groups: 17-19, 19-24, and 24 or older (when they got their licence).

Participants’ education (they were given 8 different education options and a ‘Different, namely...’ option) was dichotomized into lower education and higher education. People who filled in the ‘Different, namely...’ option were manually divided in this new variable. 11 of them were not specific enough, and were not included in analyses about education level.

Participants were asked to give an estimation of how many kilometres they had driven in the first, an in the most recent year they had had their licence. Answers of 120,000 km or higher were marked as missing because they were very improbable (10,000 or more km each month).

Outliers on driving lessons were removed (marked as missing), so no values over 200 remained. Which is still very high, but present in both the simulator group as the control group and not impossible. If the number of on-road driving lessons was 10 or lower, it was also marked as missing because that would be very unlikely and is possibly a typo.

Questions were asked about if and how many (severe) accidents participants had been involved within the first year of having a driving licence, and in the most recent year of having a licence. In analyses of accidents in the first year, only participants who had had their licence for at least a year were included; in analyses of accidents in the most recent year only participants who had had their licence for at least 2 years were included (to make sure these different periods did not overlap).

<table>
<thead>
<tr>
<th>Table 1 – Summery of subject data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age when obtaining the driving licence</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
### Results

Only the results about the transfer on the first driver exam and the retention of safe driving skills are presented.

Students who followed simulator training (8 lessons and more) with automatic feedback on viewing skills performed significantly better on the first driver exam than simulator students who did not receive automated feedback on viewing skills (7 lessons and less). Simulator students who followed simulator training without automatic feedback on viewing skills had an average passing rate on their first exam of 51.8%. Simulator students who followed simulator training with automatic feedback on viewing skills had an average passing rate on their first exam of 59.6%. In combination with a hazard perception training, the passing rate was 81.8%, 33.8% higher than the Dutch national average of 48% (over the period between 2008 and 2015).

No significant effect on accident involvement was found for simulator students who only followed vehicle handling training. Simulator students who followed driving lessons (7 or less) without automated feedback on viewing skills had an average accident involvement of 13.8% in the first 12 months after licensing. Simulator students who followed driving lessons (8 or more) on a simulator with automated feedback on viewing skills had an average accident involvement of 5.1% in the first 12 months after licensing, 32% below the Dutch national average of 7.5%. The accident involvement of the control group of students who only followed driving lessons on road was 9.9% in the first 12 months after licensing. The control group drove 2,000 km in the first 12 months. Simulator students drove 1,500 km. After correction for the exposure, the risk of being involved in an accident for simulator students who followed driving lessons on a simulator with automated feedback was 31% lower than for students who only followed driving lessons on road.

Significant differences in accident involvement were also found between gender, age groups, education types, driving styles, and learning styles.
6 Conclusions

We assumed automated feedback on viewing skills during driver education could have positive effects on the transfer and the retention of safety related driving skills and lowered the effect of EOS. We developed and implemented view assessment technology in a simulator curriculum and added a hazard perception training. Our research showed significant differences in passing rates and accident involvement between (former) students who received automated feedback on their viewing skills and (for) students who did not get this feedback during training. Students who followed driving lessons with automated feedback on viewing skills had higher passing rates on the first driving exam and needed less lessons in total. Simulator students who got automated feedback on viewing skills were less involved in accidents than simulator students who did not get this feedback and students who only followed driving lessons on road. These transfer and retention differences had a positive correlation with the amount of simulator lessons followed by the (former) students. The more simulator lessons followed with automated feedback on viewing skills, the bigger the differences between the groups. These effects clearly indicate a positive effect of our new learning theory Mental Transition on transfer and retention of safe driving skills. However, influence of self-selection, self-reporting, and non-randomised testing should be taken into account.

7 Discussion

Our research showed passing the first driver exam does not have to lead to a higher risk of accident involvement. We suggest using automated feedback on driving skills in general, and viewing skills specifically, during driver training to lower accident involvement after licensing. Driving simulators offers several advantages in comparison with cars and human supervisors, like secure safety, easy data acquisition, uniform training and assessment, lower costs for students, higher profits for driving schools and no CO2 production. We also suggest to use automated feedback after licensing to decrease EOS and increase road safety. This could be done in car but also using e-learning with driving simulation. Automated feedback on manoeuvring skills during driving on roads could also lower erosion of driving skills due to automation with ADAS and use of self-driving vehicles (Kuipers 2014).

However, we also advise to conduct a retest in a more controlled environment with randomised, filtered subjects to validate our results.

8 Acknowledgement

This research is supported by the Dutch Ministry of Economic Affairs, Green Dino BV, T. ter Morschhuizen, and J. De Winter.

REFERENCES


Automated feedback on viewing skills lowers accident involvement

Victoria, Australia: Royal Automobile Club of Victoria.


Automated feedback on viewing skills lowers accident involvement
License to Supervise. Influence of Driving Automation on Driver Licensing.

Arie P. van den Beukel, Human Centered Design Group, University of Twente, The Netherlands, a.p.vandenbeukel@utwente.nl, Chris Huijboom, Automotive Research HAN University, The Netherlands, Tobias Heffelaar, Noldus Information Technology, The Netherlands.

ABSTRACT

To use highly automated vehicles while a driver remains responsible for safe driving, places new – yet demanding, requirements on the human operator. This is because the automation creates a gap between drivers’ responsibility and the human capabilities to take responsibility, especially for unexpected or time-critical transitions of control. This gap is not being addressed by current practises of driver licensing. Based on literature review, this research collects drivers’ requirements to enable safe transitions in control attuned to human capabilities. This knowledge is intended to help system developers and authorities to identify the requirements on human operators to (re)take responsibility for safe driving after automation.

Keywords: Automation, Driving Performance, Supervision, Transition, Workload.

1 INTRODUCTION

At an increasing pace automotive industry is introducing vehicles that allow automatically driving (Broggi, Zelinsky, Özugüler, & Laugier, 2016) (Shladover, 2017). This development receives considerable attention from authorities and policy makers (Anderson et al., 2016) (Kyriakidis, Happee, & de Winter, 2015), because it may resolve mobility problems: Driving automation could increase traffic efficiency and reduce road accidents (Hoeger et al., 2008). Automated vehicles may also reduce the ecological impact of road transport (Wadud, MacKenzie, & Leiby, 2016), increase the access to mobility for elderly or disabled persons and reduce urban needs for parking (Litman, 2014).

Basic building blocks for longitudinal and lateral control have been available for years in advanced driver assistance systems (ADAS). However, safe implementation of complete automation requires highly reliable machine-based sensing of the environment in combination with faultless understanding of its environment and decision-making. Due to the highly complex traffic circumstances, current systems do not meet these requirements (Butmee & Lansdown, 2017) (Bengler et al., 2014) – leaving an important task at the driver to supervise safe system operation. Implementation of automated driving is therefore targeted from levels of so called ‘semi’ or ‘partial’ automation to ‘high’ automation – but not (yet) full automation (Gasser & Westhoff, 2012) (McGehee et al., 2016) (Shladover, 2017). This implementation means that the driver remains responsible.

For the foreseeable future, automated vehicle technologies, will continue to rely on a “responsible” driver to oversee the technology, capable of resuming control and having the foresight to make many (yet to be defined) strategic operational decisions. However, this responsibility in combination with reduced involvement in the control task causes a changed, yet difficult, role for the driver. That is: The driver’s task changes to a supervisory role during automation with the necessity to retake control during transition from automation to self-driving.
License to Supervise. Influence of Driving Automation on Driver’s License.

(Merat & Lee, 2012). However, this new role is not something humans are good at. Supervision, for instance, is accompanied with low vigilance and behavioural adaptation (Merat, Jamson, Lai, Daly, & Carsten, 2014), causing e.g. slower reaction times, misinterpretation (Martens et al., 2008) (Jamson, Merat, Carsten, & Lai, 2013) or skill degradation (Toffetti et al., 2009). Intervention is impaired by reduced Situation Awareness and increased Workload. Studies revealed that it takes considerable time to prepare for safe take-over when not being actively involved in the driving task (Radlmayr, Gold, Lorenz, Farid, & Bengler, 2014) (Gold, Damböck, Lorenz, & Bengler, 2013) (Zeeb, Buchner, & Schrauf, 2015). To summarize, partial automation creates a gap between drivers’ responsibilities and human capabilities for safe vehicle operation when using driving automation. This gap is not being addressed by current practises of driver licensing.

Assessment of driver’s proficiency (i.e. driver licensing) is essentially based on assuming a driver who is actively controlling the vehicle with minor assistance from machine-controlled functions. However, the changing role of the driver may require additional criteria to be assessed. Examples are proficiency in system supervision and the ability to safely retake control (Van den Beukel, 2016). Authorities for driver licensing in the Netherlands basically assess control-output irrespective of any assistance functions being used. For assessment of driver’s proficiency a main concept employed is: “keeping a safe ‘Spacial Cushion’” (CBR, 2008). Such Spacial Cushion is a trade-off from speed, distance (both lateral and longitudinal) and anticipation. The aim of this research is to investigate if assessment of keeping a safe Spacial Cushion can also be applied for the assessment of the driver’s supervisory role. Good anticipation is closely related to Situational Awareness (SA), especially Level 3 SA "Projection of future state" (Endsley, 1996). Situation Awareness, Distance and Speed, have been studied in many experiments in order to assess their influence on take-over behaviour and safe driving of automated vehicles. Therefore this research conducts a literature review to collect evidence whether the three aspects of a Spacial Cushion (i.e. Distance, Speed and SA Level 3) allow to univocally assess the driver’s new role. The results are intended to help system developers and authorities to identify the requirements that enable human operators to remain responsible for safe driving when operating automated vehicles.

2 METHOD

Our literature review is based on a narrative review of quantitative evidence of the parameters that define driver’s proficiency to retake control after automated driving. We defined automated driving as any system in which the automation has taken over both longitudinal and lateral control during road travelling. Articles were collected based on a combination of search items. Studies were selected for closer consideration based upon specific criteria, resulting in 68 articles. This approach is explained below.

Our literature search is conducted with Google Scholar, because this search engine has broad coverage and features full-text search. To select studies that involve driving automation, searches were based on any of the keywords “driving automation”, “automated driving”, “autonomous driving”, “driverless”, “robotic cars” or “self-driving”. These words were combined with “transition”, “take-over”, “workload”, “driver control”, “situation awareness”, “accident avoidance”, “secondary task”, “eye-tracking” and “hands-on-the-wheel”. We selected studies from 2006 onwards. Studies solely focusing on automation of exceptional manoeuvres (like parallel parking) were excluded.
License to Supervise. Influence of Driving Automation on Driver’s License.

We selected studies for closer consideration if at least one of the following aspects were being assessed:

1. Reaction time after a take-over request;
2. Situation Awareness;
3. Hazard detection;
4. Driver response to solve (potentially) critical traffic situations;
5. Performance level of non-driving tasks;
6. Eye-movement;
7. Hand position (on/off steering wheel);
8. Occurrence of driver’s initiated (de)activation of automation.

The ability to take-over control is influenced by timing aspects, i.e. the urgency of an event and driver’s reaction time (1). To assess the effectiveness of retaking control the driver’s Situation Awareness (2) is an informative measure. Based on Endsley (1996) Situation Awareness (SA) is defined as “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future”. A variety of methods exist to measure SA. We selected only studies that used Freeze Probe techniques, like SAGAT (Jones & Endsley, 2004). This is because Freeze Probe techniques show best face validity and most reliable results compared to other SA-measures (Salmon et al., 2009) (Van den Beukel & Van der Voort, 2017). Hazard detection (3) is included in the first two levels of SA, i.e. perception and understanding. We also included studies that measured hazard detection separately.

With regard to driver response (4), we included studies that measured brake-action, steering-action to avoid danger, and combinations of braking and steering. To assess performance levels of non-driving tasks (5) we included studies that quantified the involvement in non-driving tasks. Performance of non-driving tasks may also be associated with hand-position on or off the steering wheel (7). Measurement of eye-movement (6) allows to assess driver’s awareness of critical situations. Eye gaze patterns are useful for inferring workload, with reduced variance of gaze serving as an indicator of higher workload (Victor, Harbluk, & Engström, 2005). Drivers’ initiatives for (de)activation of the automation (8) is indicative for their involvement in the driving task and their anticipation of system boundaries.

For above selected measurements both driving simulator studies and studies in real cars are eligible. However, our review does only account for studies in which transitions take place. Studies were considered if published as journal article, book chapter, conference proceedings, doctoral theses or reports from research projects. To retrieve the complete manuscripts we used in addition to Google Scholar also Web of Science, Scopus and ResearchGate. For reasons of practicality publication language was restricted to English, German or Dutch. In total, 68 articles were selected to include in our review. However, this number does not yet include review of the reference lists of the selected studies in order to consider further relevant studies.

3 RESULTS

Most studies measured take-over behaviour either after a take-over request, or after a critical event taking place, requiring the driver to intervene in order to avoid an accident.
License to Supervise. Influence of Driving Automation on Driver’s License.

Reaction Time and Driver Response

Louw, Merat, and Jamson (2015) assessed Reaction Time and both lateral and longitudinal acceleration after a take-over event occurred while driving automatically. The study showed that automated driving caused longer reaction times and more agitated reactions witnessed by stronger lane deviations and more severe braking reactions (Louw et al., 2015). Merat and Jamson (2009) researched if full automation influences drivers’ response after experiencing three different events that required an intervention. In all three situations Reaction Time was faster for manual drivers than drivers of an automated vehicle. Feldhütter, Gold, Schneider, and Bengler (2017) assessed Reaction Time, Take-over Time, maximal longitudinal and lateral deceleration, Gaze behaviour and Time-to-Collision while participants took up a secondary task. During secondary task uptake, Reaction Time and Take-over Time were longer indicating that it takes additional time to switch attention to the road. Without secondary task, participants monitored the road more frequently, but during prolonged driving the monitoring time reduced.

In addition to Reaction Time Merat et al. (2014) researched time needed to obtain stabilized control of the vehicle after automation. According to their study 35-40 seconds were required for the human driver to achieve stabilised lateral control irrespective of whether handover from machine control had been planned or was in response to a critical event. Gold et al. (2013) researched Take-over Times and demonstrated that the time required to efficiently regain control in response to a critical event was increased between 2.1 to 2.89 seconds compared to response times while manually driving. Under traditional manually driving conditions a basic advise is to have at least 2 seconds time-difference with vehicles in front in order to accommodate delayed response actions due to reaction times. Applying the finding of studies like Gold et al. (2013) would mean that drivers of automated vehicles are advises to keep at least 4 seconds time-difference with a vehicle in front.

Driving Performance: Speed and Distance

Some studies found no adverse influence on Reaction Times, but impaired driving performance after retaking control. Despite immediate response to an handover signal, Merat and Jamson (2009) found that drivers braking response in an automated condition was not as effective as in the manual condition. Most likely this was due to reduced situational awareness. Also Zeeb et al. (2015) found no influence of distraction on reaction time. However the take-over quality in terms of lateral disturbance in lane position was larger for distracted drivers. These findings seem to indicate that establishing motor readiness may be carried out almost reflexively, but cognitive processing of the situation is impaired by driver distraction. According to (Gold et al., 2013) the more time given, the longer it takes the driver to intervene. If less time is available, the more disturbance in obtaining stabilized vehicle control occurs. This fact may emphasize the importance of driver’s understanding of the system’s boundaries. If drivers try to make full use of the given time budget, they need to estimate the remaining time. Thereby factors like the driver’s sensitivity to the remaining time budget, or the perceived criticality and urgency of the situation comes into play. Conditions that may impact performance include the number and type of critical incidents, traffic density, feedback, distraction, and fatigue (Merat & de Waard, 2014).
Situation Awareness and Attention

Miller et al. (2015) assumed that drivers’ attention would benefit from a sufficient level of arousal. To have sufficient arousal (not necessarily for driving monitoring but more in general) participants were therefore purposely invited to engage in secondary tasks. This was compared with drivers’ road monitoring in situations with low arousal (i.e. drowsiness). From Miller’s study it may concluded that using secondary tasks to have drivers available at a sufficient attention level causes contra productive effects: The secondary task uptake either keeps drivers alert but unaware or without such task drivers may be aware of driving but their attention is weakened due to being bored. Equally, Jamson et al. (2013) revealed that drivers in an automated condition showed 2% higher drowsiness than in manual conditions and confirmed the adverse influence of fatigue and task underload on drivers’ performance for supervising automated vehicles. Fletcher and Zelinsky (2009) assessed the driver’s monitoring role based on comparison between relevant traffic events and driver’s gaze direction. If the gaze-direction did not correspond with an event this was an indication of inattention to the event. However, the opposite cannot be concluded due to ‘look-but-did-not-see’ problems: If the gaze direction corresponded with an event drivers were nonetheless regularly inattentive.

Behavioral adaptation

Jamson et al. (2013) also revealed that drivers adapt their supervising performance to traffic circumstances. In heavy traffic conditions drivers gave more attention to the supervision task in comparison to conditions with low traffic density. With higher automation drivers are also less wanting to overtake (resulting in longer travel time). While drivers who experience automation tended to refrain from behaviours that required them to temporarily retake manual control (such as overtaking), drivers were significant less moment exposed to low time-to-collision than with manual driving. Thus far, automation improved safety margins in car following, however this was restricted to increased journey and higher involvement of drivers with in-vehicle entertainment tasks than they were in manual driving, affording less visual attention to the road ahead. This might suggest that drivers are happy to forgo their supervisory responsibilities in preference of a more entertaining highly-automated drive. However, they did demonstrate additional attention to the roadway in heavy traffic, implying that these responsibilities are taken more seriously as the supervisory demand of vehicle automation increases.

4 CONCLUDING REMARKS

This study reviewed if the three aspects contained in the concept of a Spacial Cushion, i.e. longitudinal and lateral Distance, Speed and SA Level 3 would allow for assessment of the driver’s supervisory role. The conclusion from this review is that if such assessment is applied to human-machine cooperation, it will not provide a sufficient and unambiguous picture. This is because this assessment cause fundamental contradictions with regard to distance and anticipation. That is, machine controlled functions may allow shorter time-distances due to faster reaction times. However, human fall back requires time-distances to be even longer than during manual operation. Our review showed considerable empirical evidence that drivers need additional time to get back in the control-loop after automation. Based on several studies, the additional response time
License to Supervise. Influence of Driving Automation on Driver’s License.

requires at least 2 seconds (and considerably more when drivers are completely disengaged from driving). This means that taking the traditional Spacial Cushion as a reference for the assessment of drivers of automated vehicles is an insufficient parameter for measuring supervisory performance. Furthermore, the concept of a Spacial Cushion may overlook driver’s understanding of system boundaries and anticipation to potential changes in the machine’s operational field. To assess the driver’s role in automated driving, it is therefore recommended to renew the definition of retaining a safe Spacial Cushion and to add assessment of driver’s perception and understanding of system performance (i.e. System Awareness Level 1 and 2).

Studies reviewed in this research were based on systems that required driver’s responsibility to comply with speed limits. Drivers also needed to resume manual control if they wanted to overtake a vehicle. This may have positive influences on traffic safety as drivers seem to prefer a steady traffic flow with constant speed over manual control if this is required to overtake vehicles. However, while technology advances automated take-over manoeuvres become available. Further research on the consequences if speed and distance choices in automated vehicles come under shared responsibilities between driver and vehicle is strongly needed.

Results of this literature review are intended to help system developers and authorities to identify requirements that enable human operators to remain responsible for safe driving when operating automated vehicles. The outcome of this research may have a large and diverse impact. First of all, the research helps developers and authorities (e.g. authorities for driver licensing or infrastructural planning) by improving their understanding of the limitations of automated driving systems with respect to human capabilities for retaking control. Furthermore, it supports conclusions whether operation of automated vehicles is sufficiently backed up by current levels of required driving proficiency (i.e. driving license criteria). Further research is needed to identify if additional drivers’ training is recommendable to operate automated vehicles. More generally, the results focus both on (new) criteria for drivers’ proficiency as well as criteria for systems’ conditions in order to assume safe operation. These findings could influence allowable operating range for automated driving in terms of duration, road type, traffic situations and driving conditions.

REFERENCES


License to Supervise. Influence of Driving Automation on Driver’s License.


Fletcher, L., & Zelinsky, A. (2009). Driver inattention detection based on eye gaze—Road event correlation. The international journal of robotics research, 28(6), 774-801.


Louw, T., Merat, N., & Jamson, H. (2015). Engaging with Highly Automated Driving: To be or Not to be in the Loop?


License to Supervise. Influence of Driving Automation on Driver’s License.


DRIVER STATE AND WORKLOAD
The assessment of hazard awareness skills among light rail drivers

Avinoam Borowsky, Ben-Gurion University of the Negev (BGU), Israel, borowsky@bgu.ac.il, David Shinar, BGU, Israel, Yisrael Parmet, BGU, Israel, Netta Palacci, BGU, Israel, Moshe Yitzhaki, JMTM, Israel.

ABSTRACT

Light rail (LR) is a popular means of public transportation worldwide, in use in more than 380 cities worldwide. Light Rail Drivers (LRDs) must have good hazard awareness: the ability to understand the complexity of the traffic environment and anticipate road events. Yet, no study has examined LRDs ability to anticipate hazards, and this is the purpose of this study. The experimental group included 28 certified LRDs from the LR in Jerusalem. The control group included 26 students from BGU, with no experience with LR driving. Participants observed 18 short video clips of the driver's field of view and had to press a response button each time they identified a hazard. Participants' eye movements and button presses were recorded throughout the experiment. In general, LRDs were better at identifying hazards compared to the control group. Novice LRDs with less than 1 year of LR driving experience were more likely to respond to hidden hazards. The implications are discussed.

Keywords: Light rail drivers, hazard awareness, eye movements, Light rail driving experience.

1 INTRODUCTION

Light rail (LR) is one of the most popular means of public transportation worldwide, currently operated in more than 380 cities all around the world, transferring approximately 13.6 billion passengers daily (Dauby, 2015). LR driving is a complex task that needs to be performed in a dynamic and complex environment of urban traffic, often under high levels of mental workload. According to Naweed and Rose (2015) the most risky areas with high potential for a crash include intersections that are used by all road users or areas where the track runs along the road without any segregation, where the potential for conflicts with other road users is highest and often unpredictable. While the LRD needs to safely navigate through this complex traffic environment and anticipate road hazards there are additional challenges imposed on the driver, he also has to maintain a high level of customer service by following a strict time schedule and provide a pleasant trip to the passengers (Naweed and Rose, 2015; Nazning, Currie and Logan, 2017).

According to Naweed et al. 2017, good LR driving not only requires the awareness of the various elements in the environment but also the ability to understand the environment's complexity and be prepared for unexpected events. This ability of anticipating hazardous events, known as hazard perception (HP), has been studied widely in the domain of automobile driving (e.g., Chapman and Underwood, 1998; Horswill and McKenna, 2004; Sagberg and Bjørnskau, 2006; Borowsky, Shinar, and Oron-Glad, 2010). HP can be defined as drivers' ability to "read" the road and anticipate hazardous situation (Horswill and McKenna, 2004). It is argued that of the many driving skills that a driver possesses only HP has been found to correlate with traffic crashes (Horswill and McKenna, 2004). For example, Horswill et al. (2015) found that drivers who failed in the Queensland's official hazard perception test (HPT) were 25% more likely to be involved in a crash in the preceding year as well as in the year following the test compared to drivers who passed the test successfully.
Assessment of hazard perception skills among light rail drivers

Such findings, among others, promoted HPT to be an integral part of the official licensing procedure in the UK among others since 2002 (Crundall, 2016). While HPTs have been widely studied with respect to car drivers, there seems to be no indication for using this kind of measure to evaluate LRD driving performance.

One of the most consistent findings with respect to HP is that experienced drivers possess better HP skills compared to young-novice drivers (e.g., Sagberg and Bjørnskau, 2006; Borowsky, Shinar, and Oron-Gilad, 2010). This superiority is typically reflected by faster response times to hazards (Horswill and McKenna, 2004; although see Borowsky and Oron-Gilad (2013) for further discussion on this measure), and more effective road scanning patterns (Chapman and Underwood, 1998). In addition, experienced drivers are better than young-novice drivers at anticipating hidden hazards (Borowsky et al., 2010; Vlakveld et al., 2011), situations where the hazard instigator is obscured behind an object or another road user (e.g., a pedestrian is obscured behind a parked truck and might burst into the road).

2 OBJECTIVES

The aim of this study was to generate a valid HPT for LRDs in order to evaluate their HP performance and to define a gold standard for adequate performance that can later be used by LRD trainers and other stakeholders.

3 METHOD

3.1 Participants

Fifty-four participants took part in this study. The experimental group consisted of 28 participants (one female) who were learners or novice drivers (3) or qualified LR drivers (24) of the Jerusalem LR Transport (JLRT), ranging in age from 26 to 62. The control group included 26 students ranging in age from 25 to 28. This group had no prior knowledge or experience in LR driving. All participants had at least 5 years of driving experience, uncorrected Snellen visual acuity of 6/9 (20/30) or better, and normal contrast sensitivity. All had 5+ years of driving experience. All participants filled a demographic questionnaire before the experiment began.

3.2 Materials and Apparatus

3.2.1 Eye tracking laboratory

The experiment was conducted in the eye tracking laboratory at BGU. A 20” LCD wide screen with 1360* 768 pixels, connected to a Pentium 4 PC, was used to display the movies. Participants sat 65 cm from the LCD, which provided them with a visual field of 22 degrees vertically, and 35 degrees horizontally. Another PC was used to operate the eye tracking software interface and to control the participant’s computer. Eye movements were recorded with an Eye Tracking System (ETS; Applied System Laboratories, Model D6), sampling the visual gaze at 60 Hz, with a nominal accuracy of 0.5 degrees of visual angle. The D6 facial recognition algorithm allows head free eye tracking without putting any equipment on the participant.

3.2.2 Hazard perception movies and software.

A Panasonic Lumix G DMC-G7 camera with an Olympus 7-14 mm wide lens mounted in the middle of the internal side of the driver’s cabin’s windshield directed toward the rails was used to record the video footage.
Assessment of hazard perception skills among light rail drivers

Recordings of the drive from both directions of the route were made on different days and hours with different drivers. Each recording lasted about 50 minutes and covered the whole 13.8 km route along one travel direction. The movies were edited into dozens of short segments (24-42 seconds). A representative sample of these segments were presented to a group of experts (LR drivers qualified trainers) at the JLRT in order to get a grasp of the typical hazardous situations that LR drivers might experience when driving along the route at different days and hours. Based on their comments, 18 short driving movies were selected for the experiment. Of these 18 movies, 2 were used for practice purposes only. Of the remaining 16, 13 were day-time and 3 were night time drives. Each movie contained a different number of hazardous events, totalling 50.

Movies were displayed randomly to the participant on a 20" LCD and he or she was asked to press a designated response button whenever they identified a hazardous situation. The participant's response had no effect on the movie and it continued playing to its end. At the end of each movie the participant indicated the reason for each response and typed it into a designated text-box on the screen. A fixation screen (a grey screen with a small black circle in the middle) was presented for 500msec before the beginning of the next movie. An in-house software was used to present the movies and to record all button presses and the associated reasons as well as eye movements' data.

3.2.3 LR road hazards classification

Typical hazards that appeared in the movies included situations such as pedestrians walking in parallel and near the tracks, pedestrians and bicycles who were crossing the tracks, a passing train in the opposite tracks, other vehicles crossing at crossroads, etc. Figure 1 contains examples for typical hazards.

![Figure 1 - Two examples of typical hazards in night-time (left) and daytime (right)](image)

Aiding the experts' comments, all hazards were predefined by the research group and in case of disagreements they were settled through discussion. The durations of the hazardous situations varied from 24 to 42 seconds. They were classified according to a taxonomy proposed by Borowsky and Oron-Gilad (2013) as belonging to one of three types: (1) **Visible-Potential (VP)** hazard. This type of hazard includes situations where the hazard instigator is visible, but the hazard instigator had not entered the driver's path. This type of hazard requires the driver to monitor the hazard, but it does not require any evasive action. A potential hazard can materialize at any given moment. For example, a pedestrian walking parallel to the tracks. (2) **Hidden-Potential (HP)**. This type of hazard includes situations where the hazard instigator is hidden behind other road users or behind some objects and therefore it cannot be seen by the LR driver. In this type of hazard, the hazard instigator does not enter the driver's path and the driver should monitor the area from where a hazard instigator might appear. For example, a passing train in the counter-tracks that can obscure potential pedestrians behind it. (3) **Visible-Materialized (VM)** hazard. This type of hazard is similar to visible potential hazards, but in this case it actually materializes. For instance, a pedestrian that darts into the tracks very close to the front of the train.
3.3 Procedure

The participant arrived to the lab, signed a consent form and went through a visual acuity test and a contrast sensitivity test. The participant was then asked to seat in front of the display and to fill a demographics questionnaire. Once finished, (s)he went through a short eye calibration process and a detailed instructions page was presented on the screen explaining that (s)he is about to observe a sequence of LR driving movies that were taken from a LRD’s perspective and that their task is to press a response button whenever they identifies a hazard. Next, two movies were presented as practice to allow the participant to get familiarized with the experimental task. Then, the participant was asked to start the HPT. As described above, in this phase 16 driving scene movies were displayed to the participant and he was asked to press a response button whenever he identified a hazard. At the end of each movie the participant indicated the reasons for each button press. The experiment ended at the end of this phase.

4 RESULTS

The data collected on each participant includes two types of information: (1) behavioural response, and (2) eye movements. For the purpose of this paper, we will focus on the button press analyses.

3.4 Hazard detection (probability to identify a hazard)

This measurement describes the probability of a participant to respond to a certain hazard. The dependent variable is binary distributed indicating whether a participant identified the hazard ("1") or not ("0") within the allotted time. For this analysis two independent variables were included in the model: Group (control, novice LR drivers, experienced LR drivers), and Hazard Type (materialized, potential-visible, potential-hidden). A binary logistic regression model was utilized within the framework of General Linear Mixed Models (GLMM), with a logit link function. The independent variables and their interaction were included as fixed effects and participants were included as a random effect. Applying a backwards elimination procedure, the final model’s results are summarized in Table 1. Post hoc pairwise comparisons was done using the Bonferroni correction.

<table>
<thead>
<tr>
<th>Source</th>
<th>F</th>
<th>DF1</th>
<th>DF2</th>
<th>Sig</th>
<th>Estimated Means (SE)</th>
</tr>
</thead>
</table>
| Group                   | 12.20| 2   | 2691| <0.01 | C=0.3 (0.03), NLR= 0.72 (0.07), ELR=0.44 (0.04)  
NLR >ELR >C (P_{adj}=0.01; P_{adj}=0.03 respectively) |
| Hazard Type             | 52.01| 2   | 2691| <0.01 | VP=0.47 (0.03), HP=0.26 (0.04), VM=0.74 (0.04) 
MV>VP>HP (P_{adj}<0.01; P_{adj}<0.01 respectively) |
| Group*HazardType        | 14.58| 4   | 2691| <0.01 | See Figure 2 left panel               |

Note for Tables 1 and 2. C, LRN and LRE stands for Control, Novice LR drivers, and Experienced LR drivers. VP, HP and VM stands for Visual-Potential, Hidden-Potential and Visual-Materialized hazards respectively.

Figure 2 left panel describes the interaction between group and hazard type. Post hoc pairwise comparisons analysis revealed several patterns. First, novice LR drivers and experienced LR drivers were significantly more
likely to respond to visible-potential hazards than the control group (NLR: \( P_{\text{adj}}<0.01 \), ELR: \( P_{\text{adj}}<0.01 \)). Novice LR drivers were significantly more likely to respond to hidden-potential hazards than experienced LR drivers. (\( P_{\text{adj}}<0.01 \)) and the control group (\( P_{\text{adj}}<0.01 \)). In addition, experienced LR drivers were more likely than the Control to respond to these type of hazards (\( P_{\text{adj}}=0.01 \)). Second, while Novice LR drivers had a similar response probability for all types of hazards, both the control group and Experienced LR drivers were more likely to respond to visible-materialized hazards, followed by visible potential hazards followed by hidden potential hazards (ELR: VM> VP; \( P_{\text{adj}}<0.01 \), VP> HP; \( P_{\text{adj}}<0.01 \); Control: VM> VP; \( P_{\text{adj}}<0.01 \), VP>HP; \( P_{\text{adj}}<0.01 \)).

**Figure 2 - Interaction between Group and Hazard type.**

The left panel refers to the estimated hazard probability of identifying a hazard and the right panel to the estimated medians of the normalized RT. Note: ‘*’ indicates \( P<0.05 \), ‘**’ indicates \( P<0.01 \), ns means not significant.

### 3.5 Normalized response time

This variable measures the time interval between the beginning of the hazardous situation and the time when the button press was initiated *divided by the duration of the time window of that hazard*. Since the dependent variable is bounded (between 0 and 1), it is not normally distributed. Thus, we applied a natural logarithm (LN) transformation on it. We then utilized a linear regression model within the framework of GLMM. The independent variables group and hazard type and their interaction were included as fixed effects and participants were included as a random effect. Applying a backwards elimination procedure, the final model’s results are summarized in Table 2. Estimated means were transformed back to the original values of the dependent variable (that is, before the LN transformation).

**Table 2 - A summary of the final normalized response time model’s fixed effects**

<table>
<thead>
<tr>
<th>Source</th>
<th>( F )</th>
<th>DF1</th>
<th>DF2</th>
<th>Sig</th>
<th>Estimated Medians (CI) of the normalized RTs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>6.38</td>
<td>2</td>
<td>1119</td>
<td>0.02</td>
<td>C=0.48 (0.35-0.60), NLR= 0.27 (0.12-0.48), ELR=0.67 (0.57-0.76)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ELR &gt; NLR (( P_{\text{adj}}=0.03 )), C = ELR, C=NLR</td>
</tr>
<tr>
<td>Hazard Type</td>
<td>5.58</td>
<td>2</td>
<td>1119</td>
<td>0.04</td>
<td>VP=0.58 (0.48-0.67), HP=0.34 (0.22-0.50), VM=0.48 (0.37-0.60)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VP&gt;HP (( P_{\text{adj}}&lt;0.01 )); VM=VP, VM=HP</td>
</tr>
</tbody>
</table>

**Group*HazardType 15.22 4 1119 <0.01 See Figure 2 right panel**

Figure 2 right panel describes the interaction between group and hazard type. Post hoc pairwise comparisons analysis revealed several patterns. First, experienced LRDs were significantly slower to respond to hidden-potential hazards compared to novice LRDs (\( P_{\text{adj}}<0.01 \)) and compared to the control group (\( P_{\text{adj}}<0.01 \)). No differences between the groups were found for materialized hazards. Second, while experienced LRDs
displayed significantly slower RTs for hidden hazards compared to visible hazards (HP>VP: \( P_{adj}<0.01 \); HP>VM: \( P_{adj}<0.01 \)), novice LRDs displayed an opposite pattern, that is, a faster RTs towards hidden hazards compared to both visible-potential hazards and visible-materialized hazards (HP<VP: \( P_{adj}<0.01 \); HP<VM: \( P_{adj}<0.01 \)).

5 DISCUSSION

This study examined LRDs ability to anticipate road hazards as this is one of the greatest challenges of LRDs (Naweed et al., 2017). The results of this study revealed several patterns. Because the types of hazards according which the hazards were classified were a prominent factor in affecting participants’ response the discussion will be organized accordingly. First, with respect to materialized hazards, there were no significant differences between experienced LR, novice LR and control drivers either in terms of the probability to respond or in the time to respond. This pattern is consistent with previous studies showing that even novice drivers have no problems in detecting materialized hazards and respond to them at the same speed as experienced drivers (Vlakveld et al., 2011; Borowsky et al., 2010). This pattern is not surprising because responding to a materialized hazard where the hazard instigator is visible (e.g., a pedestrian crossing the tracks in front of the LR) is independent of driving experience and does not require any anticipation capabilities. Typically, and consistent with our findings, these types of hazards induce the largest number of responses among all types of hazards.

Second, with respect to visible-potential hazards, two patterns emerged. First, experienced LRDs (0.68) and novice LRDs (0.52) were significantly more likely to respond to visible-potential hazards than the control (0.24) drivers. This is interesting because it shows that with LR training and driving experience drivers’ sensitivity to the potential hazards increases. In contrast, the control group who had no experience in LR driving was much less sensitive to visible potential hazards. Thus, an observer who is unfamiliar with the context of the LR driving environment cannot really evaluate the likelihood that a pedestrian walking along the tracks or a construction worker who stands beside the tracks will enter the LR path of travel. With accumulating LR driving experience, drivers learn that other road users in the near surroundings of the tracks may oftentimes be unexpected. For this reason, experienced LRDs tend to monitor these hazards more closely and are more sensitive to respond to these types of hazards. RT analysis showed that the control group were also much slower to respond to these hazards compared to experienced LRDs and novice LRDs (who did not differ from each other).

Third, and perhaps the most intriguing finding was the difference between experienced LRDs and novice LRDs in identifying hidden-potential hazards. It was hypothesized that experienced LRDs will identify these types of hazards better than novice LRDs. Our findings, however, revealed an opposite pattern. Novice LRDs were more likely than any other group to respond to hidden-potential hazards (0.66) followed by Experienced LRDs (0.18) followed by the control (0.09). In addition, the RT results revealed that experienced LRDs were significantly slower (MRT=0.88) to respond to hidden hazards compared to both the control (MRT=0.26) and the novice LRDs (MRT=0.06). However, given the fact that the novice LRDs consisted of only 4 people, more research should be done before drawing any conclusions.

To summarize, this study was an initial step toward establishing a HPT for LRDs; one that will hopefully be used by LR trainers as well as other stakeholders. Since the number of LRDs was limited, especially with respect to novice LRDs, more research is needed.
Assessment of hazard perception skills among light rail drivers

REFERENCES


Discriminating Drivers’ Fear and Frustration through the Dimension of Power

Meng Zhang*, meng.zhang@dlr.de, Klas Ihme*, Uwe Drewitz*

* Institute of Transportation Systems, German Aerospace Center, Braunschweig, Germany,

ABSTRACT

The goal of this study was to investigate changes in body temperature as indicators for the emotional dimension of power during driving. Therefore a driving simulator experiment with 18 participants was conducted, in which two emotions with different characteristics in the dimension of power (fear with low power, and frustration with high power) were induced using events in the driving scenarios. Changes in finger temperature, which was supposed to represent emotional dimension of power, increased significantly ($t(17) = 1.8, p < .05^*$, Cohen’s $d = 0.6$) more after frustration events than after fear. In contrast, there was no difference in skin conductance level, which indicates emotional arousal, between fear and frustration. Additionally, a preliminary analysis of face temperature points in the direction of the finger temperature. Together, the results of this study suggest that body temperature as an indicator for the emotional dimension of power can help to discriminate between fear and frustration and thus aid reliable in-vehicle emotion recognition.

Keywords: driver state, body temperature, dimension of power, frustration, fear, simulator study.

1 INTRODUCTION

Negative emotions affect humans in their different roles during highly automated driving as they impact cognitive capabilities necessary for driving, alter risk perception and influence experienced comfort and acceptance (e.g. Jeon, 2015). An automated recognition of drivers’ negative emotions could aid a solution for vehicle design because it provides the possibility to parametrize human-automation interaction according to the current emotional state of the driver. Such systems can support the human when taking over control from the vehicle or improve the comfort when the human is merely a passenger. However, automated emotion recognition is a challenging endeavour and so far no reliable methods for in-vehicle emotion recognition exist.

Most studies on automated emotion recognition assume an arousal-valence dimensional model (Russel & Barrett 1999), which considers emotion as a combination of valence (pleasure-displeasure) and arousal (calm-activity), as theoretical background. Empirical evidence suggests that these two dimensions can be discriminated using peripheral physiological measures. Skin conductance level was often used to measure the degree of emotional arousal (e.g. Lane & Nadel, 2002), while in contrast, heart rate is supposed to reflect the valence dimension (e.g. Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2000). However, changes in peripheral physiology (e.g. Skin conductance level) are not specific to emotions and vary with other driver states, such as cognitive workload (Hjortskov, Rissen, Blangsted, Fallentin, Lundberg, & Sogaard, 2004). Previous studies used the two dimensional model to measure driver emotions (e.g. Minhad, Hamid, & Reaz, 2017). Still, some emotions can barely be distinguished based on valence and arousal alone, for instance fear and frustration, both of which are unpleasant and active. However, it is meaningful to distinguish fear and frustration in driving context, since they differently impact risk perception and driving performance (Jeon, Yim, & Walker, 2011). Thus,
Discriminating Drivers’ Fear and Frustration through the Dimension of Power

A reliable assessment of driver emotions for the enhancement of human-automation interaction should not be based on peripheral physiology alone and take into account aspects of emotional experience beyond valence and arousal.

Recently, a more comprehensive dimensional model of emotions was suggested by Fontaine, Scherer, Roesch, and Ellsworth (2007). According to this model, in addition to valence and arousal, the dimensions of power and novelty play a role in describing the space of emotional experience. Here, the dimension of power represents appraisals of power or weakness feelings of control. Interestingly, this model provides evidence that fear and frustration differ on the dimension of power (fear: low power, frustration: high power). Therefore, an indicator of power may support the automated discrimination between fear and frustration. Using component analysis of phrases describing emotional experience, researchers identified “felt cold” as an indicator for a low score on the dimension of power in semantic space (Fontaine et al., 2007; Gillioz et al., 2016). Based on this, it can be assumed that changes in the dimension of power come along with changes in body temperature. Hence, the goal of this study is to investigate whether body temperature is suitable as indicator for the dimension of power to discriminate between experienced fear and frustration during driving.

2 Method

2.1 Participant

18 volunteers (four females) with an age range from 22 to 40 years (mean [M] = 27.5, standard deviation [SD] = 4.5) participated in the study. All of them possessed a valid driving license and had at least two years of driving experience. Participants provided written informed consent to take part in the study and received 15 € plus bonus depending on performance (see below) as financial reimbursement for their participation.

2.2 Set-up, Design and Procedure

The study was accomplished in a driving simulator consisting of three screens and steering wheel as well as gas and brake pedal that controlled a virtual car in a driving simulation (Virtual Test Drive, Vires, Germany). The two target emotions (fear and frustration) were introduced through different events in driving scenarios taking place in an urban setting with one lane per direction and a speed limit of 50 km/h. Participants were told that they had the task to deliver a parcel within seven minutes in order to gain a bonus of 1 € (fear) to 2 € (frustration) per drive. During the scenarios, certain events happened with the goal to induce the two target emotions (see Figure 1). Participants had to drive two scenarios per target emotion (random order) always starting with 1 min without any emotional events. After each driving scenario, participants completed self-report questionnaires on their emotional experience. The scenarios inducing the target emotions were as follows:

**Fear:** The two scenarios for fear had a length of ~5 km with three fear events per scenarios. Each fear event involved a crash or almost-crash produced by a vehicle swerving abruptly from the opposite lane (Figure 1, left).

**Frustration:** The two scenarios for frustration had a length of ~6 km creating more difficulty in the delivering task to increase participants’ motivation. Frustration was induced by blocking the road (three events per scenario), for instance through a slow lead vehicle or a traffic jam on both lanes (Figure 1, right; for similar procedure, see Ihme, Dömeland, Freese & Jipp, in press).
Discriminating Drivers’ Fear and Frustration through the Dimension of Power

Figure 1 – Screenshots of driving simulator in scenario of fear (left) and frustration (right).

2.3 Self-Report Questionnaires

After each driving scenarios, participants had to complete the Positive and Negative Affect Schedule (PANAS) and the Self-Assessment Manikin (SAM). The PANAS (German: Krohne, Egloff, Kohlmann & Tausch, 1996) is composed of 20 adjectives describing ten positive and ten negative emotions, on a Likert scale from one (low) to five (extremely). Additionally, “frustrated” was added to the list to represent frustration. We specifically focused our analysis on the items “scared” and “frustrated”. The SAM (Bradley & Lang, 1994) uses pictures to represent emotional responses on the three dimensions valence (pleasure–displeasure), arousal (excited–relaxed) and power/dominance (control–out of control). Each dimension is represented by a Likert scale from one to nine.

2.4 Physiological Measures

A finger sensor (Heally, SpaceBit, Germany) was used to assess skin conductance and finger temperature at a sampling rate of 25 Hz during the entire experiment with a sensor on the forefinger of the non-dominant hand. Skin conductance was downsampled to 10Hz, smoothed and subjected to a continuous decomposition analysis in Ledalab (Kaernbach, 2005) to separate the tonic and phasic changes. Our analysis was focused on tonic changes of skin conductance as measure of arousal. The finger temperature signal was also downsampled to 10Hz and considered as indicator for power. The signal was extracted from baseline (first ten seconds of a scenario) and an event-related epoch (from onset of events to ten seconds after).

2.5 Infrared Imaging of Faces

Participants’ faces were recorded with an infrared camera (Optris Pi640, 640*480, 10Hz) to determine variations in facial temperature from the videos. Here we present a preliminary analysis from one participant in three areas of interest (AOIs, each 5 x 5 pixels) on the face, namely forehead (between eyebrows), nose (nose tip) and cheek (centre of cheek) (cf. Merla & Romani, 2007; Pavlidis, Levine, & Baukol, 2001, see Figure 3 left). The values in the three AOIs were manually extracted using the software Optris Pi Connect (Optris, Germany).

3 Results

3.1 Manipulation Check

Participants’ rating on the PANAS item “scared” was significantly higher in the fear than in the frustration condition, while the rating on the item “frustrated” was higher (marginally significant) in the frustration compared to the fear condition (see Table 1). There was no significant difference between fear and frustration in all dimensions of the SAM (see Table 1). Still, an exploratory analysis of the PANAS provided evidence that the
Discriminating Drivers’ Fear and Frustration through the Dimension of Power

average score of a lower-order factor, which is negatively related to dominance (incl. “scared”, “nervous”, “afraid”, “guilty”, “ashamed” and “jittery”, Mehrabian, 1997), was significantly higher in fear ($M = 1.7, SD = 0.5$) than frustration ($M = 1.4, SD = 0.4$) ($t (17) = 4.5, p < .01^{**}$, Cohen’s $d = 0.03$). In contrast, another lower-order factor (incl. “distressed”, “irritable”, “hostile”, “upset”), which was solely related to valence and arousal, did not differ between the two conditions.

Table 1 – SAM and PANAS (partial) scores in fear and frustration

<table>
<thead>
<tr>
<th></th>
<th>Fear</th>
<th>Frustration</th>
<th>T-test</th>
<th>Cohen’s $d$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>PANAS-scared</td>
<td>2.7</td>
<td>1.2</td>
<td>1.6</td>
<td>0.8</td>
</tr>
<tr>
<td>PANAS-frustrated</td>
<td>2.2</td>
<td>1.0</td>
<td>2.7</td>
<td>1.3</td>
</tr>
<tr>
<td>SAM-Valence</td>
<td>5.8</td>
<td>1.9</td>
<td>5.3</td>
<td>2.0</td>
</tr>
<tr>
<td>SAM-Arousal</td>
<td>4.2</td>
<td>2.2</td>
<td>4.0</td>
<td>2.3</td>
</tr>
<tr>
<td>SAM-Dominance</td>
<td>5.5</td>
<td>1.5</td>
<td>5.5</td>
<td>1.4</td>
</tr>
</tbody>
</table>

$^T$ $p < .1$ marginally significant, ** $p < .01$ significant.

3.2 Physiological Measures

3.2.1 Skin Conductance Level

Figure 2 (left) presents the average of the changing in skin conductance level between emotional events and baseline indicating no significant difference between fear ($M = -0.1, SD = 0.9$) and frustration ($M = -0.3, SD = 0.9$) according to a T-test ($t (17) = -1.1, p = .27$, Cohen’s $d = 0.2$). This suggests little difference in arousal between the two conditions.

3.2.2 Finger Temperature

Figure 2 (right) presents the average of the changing in finger temperature, in which the difference between finger temperature in emotional events and in baseline was calculated. The result of a T-test indicates that the finger temperature increased significantly more in frustration ($M = 0.4, SD = 0.7$) than in fear ($M = 0.1, SD = 0.3$), $t (17) = 1.8, p < .05^*$, Cohen’s $d = 0.6$). This indicates that finger temperature differs between fear and frustration.

Figure 2 – Changing of skin conductance level (left) and changing of finger temperature (right)
3.3 Infrared Imaging of Faces

The face temperature analysis of the example participant indicates differences between fear and frustration (see Figure 3, right). Specifically, forehead and nose temperature increased from 5 s before the event to 10 s after the event in the frustration condition, while the opposite pattern could be observed for fear. No changes were observed for cheek temperature. Although an extensive analysis of all participants is mandatory to estimate the generalizability of these results, the revealed patterns on forehead and nose appear to be in line with the finger temperature values.

![Figure 3 – Face temperatures of one participant at three AOs (left) five seconds before, at onset and ten seconds after fear (blue) and frustration (red) events (right).](image)

4 Impact

In this study, it was revealed that body temperature can be seen as an indicator discriminating between emotional driver states with different characteristics in the dimension of power. Specifically, we could show that when experiencing an emotion with low power (fear), the finger temperature of drivers is reduced as compared to when experiencing an emotional state with high power (frustration). Interestingly, skin conductance level as measure of drivers’ emotional arousal did not differ between the two emotional states. Additionally, an exploratory analysis of a driver’s facial temperature as assessed with an infrared camera suggests that forehead and nose tip temperature appear to show a similar pattern as finger temperature. In future work, a systematic analysis of facial temperature preferably using automated measures with more emotions representing the entire valence-arousal-dominance space is needed to validate the presented results. It has to be mentioned that although the experimental manipulation seemed to be successful according to self-report, some frustration may have been unintentionally induced in the fear condition. Given that, the self-report included the complete drives, but the analysis of body temperature focussed on a small time window around the particular emotion-eliciting events, the self-report may have been biased by other factors than the mere inducing events. To sum up, drivers’ body temperature could indicate variations in the dimension of power and thus support the automated in-vehicle recognition of emotions enabling the parametrization of human-automation interaction according to the current emotional needs of the driver.
REFERENCES


Jeon, M., Yim, J. B., & Walker, B. N. (2011, November). An angry driver is not the same as a fearful driver: effects of specific negative emotions on risk perception, driving performance, and workload. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 137-142). ACM.


ABSTRACT
Understanding the impact of medical fitness to drive is important as the driving population ages. This desktop study set out to examine older driver safety from international best evidence on various aspects likely to affect an older person's fitness to drive, including the role of education, driver retraining, self-awareness, and cognitive preconditions. The review also reviewed the influence of medication and the role of the medical practitioner, as well as the effectiveness of mandatory licensing retesting. Key recommendations included the need for a standardised screening process across all Member States in assessing fitness to drive, consistent guidelines to assist medical practitioners in their role of assessing a patient’s level of safety, and promotion of materials to help older people make their own decision when to cease driving. A wider use of Medical Assessment Boards across Europe to ensure a consistent process in assessment of fitness to drive would be helpful and the development of an effective and transparent screening protocol based on functional capability is warranted when assessing fitness to drive among older drivers.

KEYWORDS: Driving, Older Drivers, Safety, Fitness, Licensing.

1. BACKGROUND
We live in an era where the population is ageing. Eurostat estimated that the percentage share of those aged 65 years and older compared with the total growth of the population in EU28 increased by approximately 8 percent between 2006 and 2016 (Eurostat 2018). In countries such as the Netherlands, Belgium, and the United Kingdom, more than 60% of the target population live in predominantly urban regions (Eurostat 2016). The UN World Population Ageing Report (UN, 2015) noted that in Europe currently, 14% of the population are aged 80 years or older and is expected to approach 30% by the year 2050.

As shown in Figure 1, Europe and North America currently lead the world in terms of the oldest aged proportions (those 80 years and older), but others (Oceania, Latin America and Asia) are not far behind and expected to catch up in the coming years. Given that fitness to drive is expected to decline as the population ages, it is critical to prepare for this changing demographic in driving and consider what needs to be done to overcome any potential increase in road trauma.

1.1 Objectives
The work reported in this paper was part of a wider study for the European Commission (DG MOVE) looking at best practice approaches to driver training, testing and medical fitness to drive (Helman et al, 2017). This part of the study set out to examine from the best international evidence available, which aspects of ageing are likely to affect an older person’s fitness to drive. Of particular importance was the safety impact of an ageing society on the driver population, the extent that unfitness to drive is a causational factor for road fatalities and serious injuries, and the identification of which mechanisms are needed to ensure safe mobility in an ageing society.

2. METHOD
The study was essentially a desktop review of best practice across a range of countries in Europe, the USA and in Australasia. Current practice across Europe was also outlined, based on the existing literature (covering the majority of Member States) where possible, and also based on responses to a short online survey (with wide participation from 25 countries in Europe).

Literature from 2000 to 2016 was sourced from international databases such as ATRI, TRID and PUBMED, using a number of pre-determined search inclusion and quality criteria around a series of research questions listed in Table 1. The search
focused on medical fitness to drive for both ageing and medical conditions that potentially put drivers at added risk of a collision. In addition, other literature known to the project team has been collected for inclusion.

Table 1: Research questions used to guide the review

- What published scientific information is available on the role of medications and reduced fitness to drive?
- Are there formal processes and requirements across Member States for assessing older and medically unfit drivers?
- What regulations and guidelines are in use in Europe and elsewhere?
- What is known about self-regulatory behaviour and compensation, and transitioning from driver to non-driver?
- What methods are available to assist self-regulation and transition and are they effective?
- What are the benefits and disbenefits of various types of licensing systems?
- What is the role of medical practitioners in the license review process in Europe?
- What is known of GPs' involvement in referrals and mandatory reporting of at-risk drivers and are these effective?
- What knowledge do GPs have regarding assessing medical fitness to drive and is there a need for guidelines and education and training?
- Do current screening tests predict poor driving ability and crash risk and are they effective?
- Is there evidence of the effectiveness of on-road tests for assessing fitness to drive?
- How medical panels are used, what is common practice, and are they effective?
- What is the role of occupational therapists (OT) and are they utilised effectively?
- Is there a need for centres of expertise for medical assessments?

The total number of articles retrieved from the search and rated by three panelists according to the quality criteria were 123 on fitness to drive (67% accepted), 62 on substance impaired drivers (32% accepted), 35 on commercial drivers’ fitness to drive (66% accepted), and 44 papers on varying topics from the TRDI/PUBMED database (73% accepted). The focus of the review was to analyse the impact of an ageing society, of their medical risks (eyesight, cardiovascular, diabetes, drugs etc.) and the likely impact on road safety. It sought to identify which medical conditions could be more prevalent with changes in the driver population and the need for additional medical checks and refresher courses.

Studies on the relationship between medical fitness to drive and ageing, evaluated against both primary and secondary outcome measures over the last 10 years. Given the large scope of this review, a series of research questions were posed for the review, as shown in Table 1 opposite. Using the evidence reviewed a series of good practice approaches were defined, and then discussed at a stakeholder workshop in September 2016, in Brussels. The focus of the workshop was on identifying barriers and enablers to implementation different good practice approaches adopted in European member countries and internationally.

3. RESULTS

3.1 Key Findings

An earlier study by Vandenberghe (2010) identified 4-key areas of importance for Europe, based on published research up until 2011. These data show that the driver licensing systems within European member states differs considerably with regard to medical fitness to drive a passenger motor vehicle (category B). While the bulk of this information is now several years old, checking from a limited sample of member-states and recent literature suggested that little has changed recently.

The overall effectiveness across the four key areas identified by Vandenberghe of (i) education, (ii) practical driver training, (iii) self-awareness of fitness to drive, and (iv) awareness of pre-conditions for safe driving shown in Table A1 in the Appendix, are still highly relevant categories. These findings have been supplemented with other more recent data and is discussed more fully in Helman et al (2016) and summarised in Table A2 in the Appendix. These findings are summarised below.

3.1.1 Education Programs

Overall, the findings from the 6-studies listed reviewed show mixed outcomes: two of these claimed improved driving at intersections and safety attitudes, while three others reported either no driving improvements and/or no crash reductions. The sixth evaluations found an actual increase in crashes for some drivers.

3.1.2 Practical Driver Training Programs

All five programs reviewed reported that practical driving training showed improvements in at least some driving skills and knowledge among older drivers. This was stronger among those tested in on-road programs although the results were not as strong for those tested in simulators. The benefit for those tested for hazard perception were effective for improving their hazard perception abilities. This suggests that such programs may have safety benefits for older drivers, although there is surprisingly little evidence attesting to their effectiveness in identifying and assessing medically at-risk drivers.

3.1.3 Self-Awareness of Fitness to Drive

Of the four studies that assessed a driver’s self-awareness of their fitness to drive, only two showed improved driving
performance. The remaining two were not particularly valid tests for awareness but more about driving quality or clinical relationships. There is a widely held assumption that older drivers have a high level of self-regulation and thought to adjust their driving behaviour to match their changing cognitive, sensory and motor capacities. It is likely that this benefit may be greater for those not suffering severe cognitive impairment.

3.1.4 Pre-Conditions for Safer Driving
Two of the three evaluation programs reviewed that sought to improve pre-conditions for driving (physical activity, speed-of-processing and reasoning) showed some driving improvements and/or crash reductions, with the other showing mixed results.

3.2 Other Related Issues of Fitness to Drive
A number of other related issues were also discussed in the Helman et al (2016) study and are reported in the following section.

3.2.1 Medication
Prescribed and over-the-counter medications play a key role in the treatment of medical conditions, short-term illness and chronic disease. Their effect and that of multiple medications (polypharmacy – the use of four or more medications) on a person’s ability to drive can be variable. Rates of per-capita prescriptions and over-the-counter medications and dietary supplements have increased considerably over the last few decades in many developed countries including Europe. It should be cautioned that an association between drugs and impaired driving does not necessarily imply causation, as other factors may be at play, such as chronic disease, acute emotional or physical stress, and performance bias. Understanding the degree of reduced driving capability and increased risk caused by medication and drug use per se presents a major challenge for road safety in separating the cause of medication influence from the underlying condition.

3.2.2 Medical Practitioner’s Role in Reporting At-Risk Older Drivers
Medically at-risk drivers come to the attention of licensing authorities primarily through referrals from a variety of sources, including physicians, law enforcement, and the court system. In most jurisdictions, referrals are also accepted from family, friends, and other concerned citizens. Common reasons for referral of older drivers include getting lost, crashes, ‘fender benders’, and ‘near misses’ associated with erratic driving and confusion.

Most medical practitioners (GPs) accept that they have a role to play in reporting those with a relevant and severe medical condition likely to affect a patient’s ability to drive, they are concerned about their abilities to make this assessment. Furthermore, most also believe that reporting of unsafe drivers to the driver licensing authority would impact negatively on the doctor–patient relationship, a concern common throughout the literature.

There is clearly a need here for guidelines and additional information to help medical practitioners in making this judgement. The findings shown in Table A2 in the appendix were mixed across member states in terms of providing such information that opens the possibility for a more European-wide program in this area.

3.2.3 Medical Advisory Boards
Medical Advisory Boards (MABs) have been established within some licensing authorities and have the responsibility of ultimately determining fitness to drive and licensure. There are various examples of MAB throughout Europe, the USA and elsewhere, however there are few evaluations regarding the effectiveness of these systems, and therefore little evidence of ‘best-practice’. MABs vary in form from adequately managing the safe mobility of at-risk older drivers, to simply offering a form of follow-up assessment when a GP reports someone of concern. The benefit of these boards is that it takes the ultimate responsibility of a person losing their license away from the GP to that of the authority and hence overcome some of the medical practitioner’s concerns in the process.

3.2.4 Mandatory license retesting
The requirement for older road users to demonstrate their continuing ability to drive has created much concern and angst worldwide amongst road safety and ageing specialists, transportation and health authorities. Supporters of the practice of periodic mandatory licence retesting for older drivers argue that people in their later years wishing to retain a licence need to demonstrate they are fit and capable of driving without increased risk to other road users. Those who oppose age-based, periodic licence retests, base their claims on the inability of licence tests to discriminate those at risk, and issues of cost-effectiveness, discrimination, equity issues, individual differences in the ageing process, and the consequence of restricting mobility based on a person’s age.
There does not appear to be a “gold-standard” approach to these assessments. On-road is commonly used for these assessments but fraught with subjectivity, validity and indeed, danger in taking a potential at-risk driver in normal traffic. Off-road testing has been an attractive alternative, yet the review failed to demonstrate an effective scientifically robust available test to date. This is something worthy of further research.

4. DISCUSSION AND IMPACT

The medical fitness to drive is still an important safety and societal issue that member states need to actively work on. The literature review identified a number of issues around the research questions listed earlier. Calls for a general screening of the whole population by age as a means of identifying at-risk drivers is unlikely to be effective, both from a performance and from a cost-benefit perspective. There is clearly an urgent need to be able to assess fitness to drive among an older person to ensure they and the rest of the population’s safety is optimized. Fortunately, older drivers tend to drive shorter distances which minimises their exposure to risk.

Importantly, drivers of all ages are unwilling to voluntary surrender their licence, and more so for the elderly as its value increases due to their reliance on cars as they lose physical abilities and become more frail. Access to private motor vehicles is critical when public transport is either unavailable locally or difficult to access and where they need for access to medical facilities increases with age.

4.1 Discussion of Good Practice

Based on the literature review and the user survey findings, several good practices were identified for medically at-risk and/or older drivers, as detailed below:

1. A European-wide consistent screening process is required to ensure a common approach for assessing at-risk drivers across the Member States. While one could argue that individual states should decide for themselves what level of risk they believe is acceptable, this will not provide optimal safety or consistency across Europe;
2. A validated off-road assessment tool is required to minimise the number of potentially at-risk drivers on the road, to ensure an improvement in safety for themselves the general public;
3. Medical practitioners (GP) are clearly a critical part of the identification of older and medically vulnerable at-risk drivers. Many stressed the lack of and need for guidelines to assist them in their assessment;
4. In addition, GPs commonly seek assistance with this task and education programs for GPs to assist them is warranted;
5. Some older drivers can make rational decisions about their own abilities to continue to drive. Materials to aide their decision in arriving at such a decision would also be very helpful; and finally
6. Some member states currently provide restricted licences to allow those at slight risk to continue to drive where public transport is limited. This practice should be consistent across all member states.

4.2 Expected Impact

From the review, a number of priority issues seemed important for consideration in Europe as listed below:

- The need for a standardised screening process across all Member States in assessing a driver’s fitness to drive is warranted, based on international best practice.
- Consistent guidelines for medical practitioners and promotion of materials to support self-regulation towards reduced driving and cessation would further help older people make the decision when to cease driving themselves.
- A wider use of Medical Assessment Boards to ensure the licensing authorities have more of a major say in removing an older person’s right to be licensed would also help to take away some of the medical practitioners’ concerns of patient blame.
- The development of an effective and transparent screening protocol for use across Europe for testing the functional capabilities of at-risk older drivers is warranted.

Acknowledgements

This study was funded by the European Commission (DG MOVE). The information and views set out in this study are those of the authors and do not necessarily reflect the official opinion of the Commission. The Commission does not guarantee the accuracy of the data included in this study. Neither the Commission nor any person acting on the Commission’s behalf may be held responsible for the use which may be made of the information contained therein.
REFERENCES


## Appendix Table A1
Evaluation studies of older driver training and education programs published 2000-early 2011.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Nature of participants</th>
<th>Age of key participants</th>
<th>Programme</th>
<th>Effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Education only</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bédard et al. (2004)*</td>
<td>‘Normal’ active drivers</td>
<td>55 yrs+</td>
<td>55-Alive/Mature Driving</td>
<td>Ineffective – no driving improvements</td>
</tr>
<tr>
<td>Nasvadi &amp; Vavrik (2007)*</td>
<td>‘Normal’ active drivers</td>
<td>55 yrs+</td>
<td>55-Alive/Mature Driving</td>
<td>Ineffective – increased crash rates for some participants</td>
</tr>
<tr>
<td>Owsley et al. (2003)*</td>
<td>Visually impaired, crash-involved drivers</td>
<td>60 yrs+</td>
<td>Tailored programme</td>
<td>Effective – self-reported improved safety attitudes, self-regulatory practices</td>
</tr>
<tr>
<td>Owsley et al. (2004)*</td>
<td>Visually impaired, crash-involved drivers</td>
<td>60 yrs+</td>
<td>Tailored programme</td>
<td>Ineffective – no crash reductions</td>
</tr>
<tr>
<td>Kelsey &amp; Janke (2005)</td>
<td>Drivers with ‘unclean’ records</td>
<td>70 yrs+</td>
<td>Education publications and/or resources list</td>
<td>Ineffective – no crash, violation reductions (although increased driving, safety knowledge)</td>
</tr>
<tr>
<td>2. Practical driver training programmes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marottoli et al. (2007b)*</td>
<td>‘Normal’ active drivers</td>
<td>70 yrs+</td>
<td>55-Alive/Mature Driving on-road training</td>
<td>Effective – improved driving knowledge, performance</td>
</tr>
<tr>
<td>Bédard et al. (2008)*</td>
<td>‘Normal’ active drivers</td>
<td>65 yrs+</td>
<td>55-Alive/Mature Driving on-road training</td>
<td>Effective – improved driving knowledge, performance</td>
</tr>
<tr>
<td>Romoser &amp; Fisher (2009)</td>
<td>‘Normal’ active drivers</td>
<td>70 yrs+</td>
<td>‘Active’ simulator training, ‘passive’ education</td>
<td>Effective only for active group—improved simulator and driving performance at intersections</td>
</tr>
<tr>
<td>3. Self-awareness of fitness to drive</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eby et al. (2003)*</td>
<td>‘Normal’ active drivers</td>
<td>65 yrs+</td>
<td>Self-awareness knowledge workbook</td>
<td>Effective – self-reported improved awareness; Valid – self-reported difficulties with driving performance</td>
</tr>
<tr>
<td>Molnar et al. (2010)</td>
<td>‘Normal’ active drivers</td>
<td>65 yrs+</td>
<td>Self-awareness and knowledge computer program</td>
<td>Effective – self-reported improved awareness; Valid – self-reported difficulties with driving performance</td>
</tr>
<tr>
<td>Bédard et al. (2011)</td>
<td>‘Normal’ active drivers</td>
<td>50 yrs+</td>
<td>Roadwise Review</td>
<td>Not valid – no association between test performance and related clinical measures, driving performance</td>
</tr>
<tr>
<td>4. Pre-conditions for safer driving</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marottoli et al. (2007a)*</td>
<td>Drivers with physical impairments</td>
<td>70 yrs+</td>
<td>Daily exercise programme</td>
<td>Effective – better driving performance than a control group</td>
</tr>
<tr>
<td>Roenker et al. (2003)*</td>
<td>Visually impaired drivers</td>
<td>48 yrs+</td>
<td>Speed-of-processing or simulator training</td>
<td>Mixed results</td>
</tr>
<tr>
<td>Ball et al. (2010)</td>
<td>‘Normal’ active drivers</td>
<td>65 yrs+</td>
<td>Speed-of-processing, reasoning or memory training</td>
<td>Effective for speed-of-processing, reasoning training – reduced at-fault crash involvement</td>
</tr>
</tbody>
</table>

Note: “‘normal’ active drivers” were a convenience samples of active older drivers NOT selected on the basis of specified medical or performance criteria.
## Appendix Table A2

**Driver Assessment procedures for some EU Member States** *(Source: Vandenberghe, 2010, and follow-up survey)*

<table>
<thead>
<tr>
<th>Member States</th>
<th>Age-base d test req’d</th>
<th>Age for first retest</th>
<th>Holder legally bound to report ill-health</th>
<th>Medical check req’d for relicence</th>
<th>GP bound to report at-risk driver</th>
<th>GP initiate need for retest</th>
<th>Medical advisor assess req’d</th>
<th>Eye-test req’d for relicence</th>
<th>On-road test req’d for relicence</th>
<th>Conditional (restricted) Licence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>yes</td>
<td>Every 10yrs</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>If req’d</td>
<td>yes</td>
<td>Unk</td>
<td>yes</td>
</tr>
<tr>
<td>Belgium</td>
<td>yes</td>
<td>50 (+5/3)</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>If req’d</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>yes</td>
<td>50 (+1)</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>If req’d</td>
<td>yes</td>
<td>yes</td>
<td>Unk</td>
</tr>
<tr>
<td>Denmark</td>
<td>yes</td>
<td>50 (+5-1)</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>If req’d</td>
<td>yes</td>
<td>Unk</td>
<td>yes</td>
</tr>
<tr>
<td>Estonia</td>
<td>yes</td>
<td>50 (+5)</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>If req’d</td>
<td>yes</td>
<td>Unk</td>
<td>Unk</td>
</tr>
<tr>
<td>Finland</td>
<td>yes</td>
<td>50 (+5)</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>If req’d</td>
<td>yes</td>
<td>Unk</td>
<td>Unk</td>
</tr>
<tr>
<td>France</td>
<td>yes</td>
<td>60 (+/5/2)</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>If req’d</td>
<td>yes</td>
<td>Unk</td>
<td>Unk</td>
</tr>
<tr>
<td>Germany</td>
<td>yes</td>
<td>50 (+5/3)</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>Unk</td>
<td>yes</td>
<td>Unk</td>
<td>Unk</td>
</tr>
<tr>
<td>Great Britain</td>
<td>yes</td>
<td>45 (+5/1)</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>If req’d</td>
<td>yes</td>
<td>yes</td>
<td>Unk</td>
</tr>
<tr>
<td>Greece</td>
<td>yes</td>
<td>65 (+5/3)</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>If req’d</td>
<td>yes</td>
<td>Unk</td>
<td>Unk</td>
</tr>
<tr>
<td>Hungary</td>
<td>yes</td>
<td>45 (+3/2)</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>If req’d</td>
<td>yes</td>
<td>Unk</td>
<td>Unk</td>
</tr>
<tr>
<td>Iceland</td>
<td>yes</td>
<td>65 (+5/4/1)</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>police</td>
<td>yes</td>
<td>Unk</td>
<td>Unk</td>
</tr>
<tr>
<td>Ireland</td>
<td>yes</td>
<td>60 (+3/1)</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>Unk</td>
<td>yes</td>
<td>Unk</td>
<td>Unk</td>
</tr>
<tr>
<td>Latvia</td>
<td>yes</td>
<td>60 (+3)</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>If req’d</td>
<td>yes</td>
<td>Unk</td>
<td>Unk</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>yes</td>
<td>50 (+5/3)</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>If req’d</td>
<td>yes</td>
<td>Unk</td>
<td>Unk</td>
</tr>
<tr>
<td>Norway</td>
<td>yes</td>
<td>60 (+5/1)</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>If req’d</td>
<td>yes</td>
<td>Unk</td>
<td>Unk</td>
</tr>
<tr>
<td>Netherlands</td>
<td>yes</td>
<td>70 (+5)</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>If req’d</td>
<td>yes</td>
<td>Unk</td>
<td>Unk</td>
</tr>
<tr>
<td>Poland</td>
<td>yes</td>
<td>55 (+5/2/1)</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>If req’d</td>
<td>yes</td>
<td>Unk</td>
<td>Unk</td>
</tr>
<tr>
<td>Portugal</td>
<td>yes</td>
<td>40 (+5/3/2)</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>If req’d</td>
<td>yes</td>
<td>Unk</td>
<td>Unk</td>
</tr>
<tr>
<td>Spain</td>
<td>yes</td>
<td>45 (+5/3/2)</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>If req’d</td>
<td>yes</td>
<td>Unk</td>
<td>Unk</td>
</tr>
<tr>
<td>Sweden</td>
<td>yes</td>
<td>45 (+10)</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>Unk</td>
<td>yes</td>
<td>Unk</td>
<td>Unk</td>
</tr>
<tr>
<td>Switzerland</td>
<td>yes</td>
<td>45 (+10)</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>Unk</td>
<td>yes</td>
<td>Unk</td>
<td>Unk</td>
</tr>
</tbody>
</table>
Drivers’ recovery performance in a critical run-off-road scenario – A driving simulator study

Alexander Eriksson, Bruno Augusto, Niklas Strand, Jesper Sandin
The Swedish National Road and Transport Research Institute, Sweden
{alexander.eriksson, Bruno.augusto, Niklas.strand, jesper.sandin}@vti.se

ABSTRACT

Single vehicle accidents are commonly caused by fatigue and distraction and resulting in severe casualties and high economic costs. In order to evaluate driver recovery from run-off-road accidents, comprising of 80% of fatal crashes on rural roads, a simulator study in an advanced full-motion driving simulator was carried out. Drivers were given a secondary task to perform at six positions down the road (to simulate distraction), and an artificial yaw deviation was added to the vehicle to induce a run-off-road accident whilst the driver was distracted. The results show that the severity of the recovery manoeuvre was larger than similar events caused by the failure of automated lane keeping systems, leading to lane departures. Furthermore, significant learning effects was found, providing recommendations for further studies into run-off-road experiments.

Keywords: run-off-road, crash scenario, driver behaviour, driver recovery, vehicle recovery.

1 INTRODUCTION

Single vehicle accidents are commonly caused by fatigue and distraction and lead to severe casualties and high economic costs. In fact, 80% of fatal crashes in the US (between 1991-2007) on rural roads were run-off-road (ROR) accidents, and rural roads are more likely to be the scene of ROR crashes than urban roads (Liu & Subramanian, 2009). Additionally, driver distraction seems to be a significant contributor to ROR accidents, as a driver is 67.1% more likely to experience a fatal ROR accident out of all types of on-road crashes (Liu & Subramanian, 2009). An analysis of 70 critical, but non-accident, run-off-road situations showed that about 70% of the drivers had been engaged in a secondary task that involves both a visual and a manual component, typically leading to situations where the driver has the left hand on the steering wheel while looking and grasping after something with the right hand (Petersson, Svanberg, & Johansson, 2013; Sandin, Augusto, Johansson, Svanberg, & Petersson, 2015). As ROR accidents are notoriously difficult to replicate, and manouvisual distraction tend to be a substantial contributor to such accidents, a simulator study was conducted to assess how drivers handle an artificially induced ROR scenario with a strong manouvisual distraction component. The experiment was carried out in a driving simulator as driver behaviour in critical ROR situations can safely be studied in advanced moving-base driving simulators. It is anticipated that the insights generated in this study may contribute to the development of advanced driver assistance systems to reduce the impact and severity of such incidents.
2 Method

2.1 Participants

12 participants (75% male, age: 33.8 ±8.9) partook in this study. All participants provided their informed consent and the study complied with the American Psychology Association’s ethics guidelines.

2.2 Apparatus

The study was conducted in the VTI simulator IV, a high-fidelity, full-motion driving simulator located in Gothenburg, Sweden. The driving simulator uses the front half of a Volvo XC60, running a validated vehicle model from a Volvo V60 and has movement in 8 degrees of freedom and 180 degree field of view. As part of this experiment drivers also had to engage in a secondary task which consisted of a small touchscreen display located to the right of the centre stack. The task was initiated by a computer-generated voice stating ‘läs siffrorna nu’ (read the numbers now, in Swedish) in the vehicle cabin, and upon hearing the sound, the participant had to place, and hold their finger on the display to display 6 numbers in quick succession and read them out loud (thus simulating a distracting event involving a visual and manual component). Each number was displayed for 200-milliseconds with a 200-millisecond interval, for a total of 2.2 seconds.

2.3 Dependent measures

The following dependent measures were used to assess driver behaviour in the critical ROR scenario:

- Lateral position in meters which provides an indication of vehicle control and indicates whether the vehicle has left the roadway or not.
- Steering wheel angle [rad] indicating controllability, and whether compensatory actions are taken by the driver prior to the distraction task.
- Vehicle heading [degrees] indicating the severity of the ROR event and recovery.
- Lateral Jerk [m/s^3] indicating the severity of the recovery manoeuvre.
- Reaction time [seconds] indicating how long drivers took to make the first manoeuvre in the recovery of the vehicle.

2.4 Analysis

The data was analysed using the Manhattan-plot technique introduced by (Gibson, 2010) and used in the human factors literature by (Eriksson et al., Accepted; Eriksson & Stanton, 2017; Petermeijer, Cieler, & de Winter, 2017)

2.5 Scenario

The participant drove a 20-minute highway scenario whilst engaging in the secondary task for a total of 6 times. On two occasions, a yaw deviation was applied to the vehicle whilst the driver was engaged in the secondary task, provoking a ROR scenario. The yaw deviation was introduced on the fourth and sixth event respectively. The yaw deviation was initiated when the driver engaged in a visoumanual secondary task on a screen. The driver had to hold their right index finger on the screen to activate the secondary task which activated an added clockwise yaw deviation, intended to create the ROR scenario. The artificial yaw deviation was included to
Drivers’ recovery performance in a critical run-off-road scenario – A driving simulator study

assure that a ROR event would occur as these events are unlikely to occur in the short amount of time a driver spends in the simulator. The added yaw deviation was presented visually when the participant engaged in the additional task but was not represented in the vehicle dynamics or the lateral acceleration of the simulator’s motion system. This was done to ensure that the participant would not notice the deviation until the vehicle had left the roadway and received the associated kinematic feedback.

2.5.1 Yaw deviation

The manoeuvre consisted of a rotation and displacement of the world in order to promote a set of vehicle states that lead to a run of road situation. Said manoeuvre was triggered when the driver activated the distraction task by pressing on the touch screen. The motion of the vehicle was characterized as a 5th polynomial describing its lateral displacement on the road. The manoeuvre was performed in open loop, i.e., the terms of the polynomial were computed when the manoeuvre initiated and remained constant until the end. The vehicle longitudinal speed was assumed to be constant and equal to the speed at the beginning of the deviation at the moment the polynomial terms are computed.

2.5.2 Manoeuvre parameters

The manoeuvre parameters were:

- \( v_x \) = variable : Vehicle longitudinal speed. Measured at the time the polynomial terms were computed. Assumed constant during the manoeuvre, i.e, \( v_x = v_{xi} = v_{xf} \).

- \( \phi_i \) = variable : Initial yaw angle towards the road edge.

- \( \phi_f = -3.0 \text{deg} \) : Final yaw angle towards the road edge.

- \( a_{yi} \) = variable : Initial lateral acceleration of the vehicle relatively to the road edge.

- \( a_{yf} = 0 \) : Final lateral acceleration of the vehicle relatively to the road edge.

- \( v_{yf} = \tan(\phi_f) v_x \) : Final lateral velocity of the vehicle relatively to the road edge.

- \( v_{yi} = \tan(\phi_i) v_x \) : Initial lateral velocity of the vehicle relatively to the road edge.

- \( y_t \) = variable : Total lateral displacement of the vehicle relatively to the road edge. Represented by the distance between the right from wheel and the road markings.

- \( d_t = 2.0 \text{sec} \) : Total duration of the artificial deviation.

Taking the displacement, position and acceleration, it was possible to compute the terms of the polynomial which characterized the vehicle lateral displacement for the duration of the manoeuvre.

3 Results

Preliminary analysis showed that all participants went off road in the first run-off road scenario. It took drivers approximately 2.78±0.34 seconds from engaging with the secondary task until peak-counter steering to bring the vehicle back onto the road (combination of reaction time, and initial correction time as per SAE J2944,
Drivers’ recovery performance in a critical run-off-road scenario – A driving simulator study

2015). It was found that the maximum lateral jerk experienced when bringing the vehicle back onto the road was $15.97 \pm 7.98 \text{ m/s}^3$ when first experiencing the run off road scenario, and $9.91 \pm 6.68 \text{ m/s}^3$ when experienced for the second time, resulting in a significant difference ($t(11) = 3.04, p<0.05$).

3.1 Learning effects

As the study entailed two ROR scenarios, learning effects were assessed. The significant difference in lateral jerk indicated a learning effect, which, combined with the significantly different lateral positioning (Figure 1) before the distraction task indicates that drivers adopted a strategy to reduce the likelihood of leaving the roadway when engaging in the secondary task by steering the vehicle slightly to the left when prompted to engage in the secondary task. This is also reflected in a reduction of vehicle heading (Figure 2) for the second ROR event, which also yielded smaller steering wheel angles (Figure 3) than the first run-off-road event.

![Figure 1](image_url)  
Figure 1. Lateral deviation (m) for the two ROR scenarios. The bottom graph shows paired T-tests and effect size comparing the two events over time. The blue vertical line indicates the start of the secondary task, and the two horizontal lines in the bottom graph indicate alpha values of 0.05 and 0.01 respectively.
Drivers’ recovery performance in a critical run-off-road scenario – A driving simulator study

Figure 2. Vehicle heading in the two ROR scenarios.

Figure 3. Steering wheel angle for the two ROR events.

4 Conclusion

This manuscript assessed how drivers handle a critical ROR scenario whilst being distracted in a full motion driving simulator. When comparing the lateral jerk values in this study with a similar event by Wörns (2018), also in VTI Simulator IV, where an active steering system failed in a left curve, the recorded maximum lateral jerk value was 6.55-14.37m/s² depending on whether the participants were fast responders or not, indicating that the participants in this study exhibited a harsher response caused by the kinematics associated with going off-road in the moving base simulator. The results presented in this manuscript serves as a baseline source of information on how drivers recover from a run off road scenario. A proposed measure to reduce the likelihood of ROR accidents could be an automatic steering-wheel intervention that in emergency situations prevents the
Drivers’ recovery performance in a critical run-off-road scenario – A driving simulator study

vehicle from leaving the lane. However, previous studies have shown that the drivers are likely to counteract such interventions so that lower level torque overlays have little effect, and higher-level overlays lead to controllability issues. Therefore, it is recommended that further studies explore the impact of active interventions. It is also concluded that experiments assessing ROR situations should be carried out in a between-group design as there is substantial learning effects resulting in compensatory strategies to avoid or reduce the criticality of such an event.

5 Impact

The scenario and the results from this study will be used as a baseline for future evaluations of safety functions that prevents the vehicle from leaving the road. Additionally, the results may be used to create driver models for how drivers handle ROR events.

6 Acknowledgement

This work was funded by the Swedish excellence centre Virtual Prototyping and Assessment by Simulation (ViP) and VINNOVA as part of the ELKA and QUADRAE projects.

7 References


Drivers’ recovery performance in a critical run-off-road scenario – A driving simulator study

Toward zero traffic accidents, 2015.

Chalmers University of Technology, Gothenburg, Sweden
POSTER PITCHES
The Organizational Response to Automation Support Degradation.  
Identifying Air Traffic Control Sources of Resilience in Cases of Radar Loss

Luca Save, Deep Blue s.r.l., Italy, luca.save@dblue.it, Daniele Ruscio, Deep Blue s.r.l., Italy, Valentina Cedrini, ENAV s.p.a., Italy, Laura Cafiero, ENAV s.p.a., Italy, Maurizio Mancini, ENAV s.p.a., Italy.

ABSTRACT
Controllers working in Area Control Centres are supported by highly automated multi-radar tracking system displaying air traffic at different working positions. The automation provides and selects for the controllers the most relevant information associated to each aircraft and assists them while they issue instructions to pilots. But, what happens when the automation fails? These rare situations represent an example of abrupt transition from a highly automated to a severely degraded mode of operation. Exploring sources of resilience during this critical loss from highly automated tasks allows to better understand the effects of automation in a complex transportation system. Safety management and business continuity considerations were made during the DARWIN project whose goal is the development of resilience management guidelines for critical infrastructures. A specific part of the DARWIN resilience management guideline was applied to a loss of radar information scenario, based on the experience of a real event occurred in a European Area Control Centre, in order to assess the resilience responses to automation degradation. The DARWIN guideline concepts supported the identification of individual and organisational resilience mechanisms that allowed the Area Control Centre to operate in a severely degraded mode, with no negative effects on the safety and a very limited impact on the business continuity. Similar adaptive capacity mechanisms can be adopted, for analogy, in other critical transportation infrastructures that aim to achieve highly automated functionalities in the management of assisted human-machine interactions.

Keywords: Automation, Resilience Management, Organizational Response, take-over control.

1 INTRODUCTION
Area Control Centres (ACCs) are responsible for controlling air traffic, from the moment an aircraft takes off from a given airport, until it lands to another airport. Differently from Tower controllers, who rely on out-of-the-window scan over the airport surface, the controllers working in Area Control Centres are always supported by highly automated systems. These mainly consist of a multi-radar tracking system displaying air traffic at different working positions, where the controllers can monitor the evolution of traffic in their respective sectors of responsibility. The automation provides and selects for the controllers the most relevant information associated to each aircraft and it assists them with a set of tools to issue instructions to pilots via radio-telephony and other communication means. But, what happens when the automation fails? Are the air traffic controllers still able to provide instructions to the pilots in case of a sudden loss of radar information? Which resilience capabilities are needed to successfully face sudden automation support degradations? Total losses of radar information are rare events and they represent a form of abrupt transition from a highly automated to a severely degraded mode of operation, in a safety-critical transportation system. They require
The Organizational Response to Automation Support Degradation. Identifying Air Traffic Control Sources of Resilience in Cases of Radar Loss

that all the Air Traffic Controllers (ATCOs) of the affected ACC handle the traffic (overflying or going to land in their area of responsibility) without being able to visualize it on their radar screens. Exploring sources of resilience in the management of this type of emergencies was one of the topics considered during the DARWIN project, whose goal is the development of resilience management guidelines for organizations managing critical infrastructures, such as the transportation ones. Traditional risk management approaches focus on prediction, prevention and protection against expected events. These approaches cover known system disturbances as initiating events. Consequently, procedures, training, and regulations for operations are put in place to protect from known disturbances and mitigate their consequences. In recent years, different methods have been proposed to overcome the limitations of these approaches, by addressing the interdependencies between critical infrastructures and the increasing complexity of the situations that can escalate off-the-scale. The guideline approach used in the DARWIN project follows this path, by proposing strategies to address the enhancement of the abilities of an organisation to sustain adaptability and continue operations as required to a changing context (Hollnagel, Woods, Leveson, 2006).

The present paper proposes the analysis of a radar loss event based on the use of the DARWIN guideline, to show how it is possible to go beyond traditional risk management practices in response to automation degradation that mainly considers just the individual take-over response (e.g. Zeeb et al., 2015). The analysis will highlight both the reaction capabilities of a single organization in case of automation degradation and the resilience management strategy of more organizations cooperating in the management of the same crisis.

2 METHOD

In order to evaluate the applicability of a specific part of the DARWIN guideline, a workshop was organized with air traffic controllers, safety experts and security experts, in order to explore the available sources of resilience in a radar loss scenario, based on the experience of a real event occurred in a European ACC.

2.1 DARWIN Concept Cards for Assessing Automation support degradation

The DARWIN Resilience Management Guideline is a WIKI based tool\(^1\) structured around a set of so-called Concept Cards (CC) that are made available to resilience practitioners. Each concept card is linked to a specific resilience management principle and suggests a number of actions that an organization managing a critical infrastructure should perform in order to increase its level of resilience. For analysing the total radar loss situation, a specific CC named “Identifying sources of resilience” was used in a pilot exercise involving representatives of different ACCs. The CC included a number of triggering questions aimed at identifying the most effective organizational response to a contingency, emergency or crisis situation, which in this case was

\(^1\) For further information, please refer to https://h2020darwin.eu/wiki/
The Organizational Response to Automation Support Degradation. Identifying Air Traffic Control Sources of Resilience in Cases of Radar Loss

generated by the degradation of an automation support system. The triggering questions were clustered in different thematic areas that were intended to encourage the practitioners to identify specific sources of resilience available in everyday operations. The thematic areas addressed for this case were: “Adaptive Capacity”, “Operational Margins”, “Resources”, while other thematic areas presented in the CC (“Monitoring”, “Goals Trade-offs”, “Dependencies and Interactions”) were not considered relevant for this scenario.

2.2 Sample

The Air Navigation Service Provider (ANSP) is the organization more impacted by a loss of radar information. Therefore, air traffic control experts from different departments of a European ANSP were involved in the pilot exercise. More precisely, the experts had working experience in three different ACCs. One of them (ACC “A”) was the control centre that experienced the radar loss, while the other two (ACC “B” and “C”) were neighbouring ACCs, i.e. control centres located in different geographical areas and managing the traffic in different airspace portions, but with at least one border shared with the ACC affected by the radar loss. The roles of the actors participating in the exercise were the following: a representative of the Security Operation Centre of the ANSP, two representatives of the ANSP Safety Department, an expert of contingency plans and four ATCOs, with working experience in all of the involved ACCs (i.e. one from ACC “A”, two from ACC “B” and one from ACC “C”).

2.3 Procedure

The evaluation session consisted of a workshop (see Figure 1) with all the experts brainstorming guided by a handout version of the card, in the identification of the available sources of resilience in a radar loss scenario. One part of the discussion focused on a BEFORE CRISIS scenario (how could we manage a radar loss situation due to a cyberattack if it occurs in one of our ACCs?). While a second part focused on an AFTER CRISIS scenario (how did we manage a radar loss situation actually occurred in one of our ACCs, due to a technical failure?). The two different perspectives were suggested by the format of the CC, which includes specific triggering questions to be used before, during or after a crisis, contingency or emergency.

Figure 1 – Evaluation session to identify the sources of resilience available in the automation degradation
The Organizational Response to Automation Support Degradation. Identifying Air Traffic Control Sources of Resilience in Cases of Radar Loss

The analysis of the before crisis situation allowed to address preparedness and all the cyber-security issues related to the possible causes of a radar loss, which in this case were never experienced by the ANSP. The analysis of the after crisis situation was equally relevant, due to the possibility to learn from a concrete experience of a radar loss event happened in 2017. The event was initially caused by a minor technical failure occurred at an airport and then propagated into unexpected cascading effects to the ACC “A”, causing the freezing of radar screens of the Controller Working Positions (CWPs) for more than two hours, during morning operations.

Following the brainstorming on the sources of resilience, a semi-structured questionnaire was also administered, in order to: a) collect feedback on the perceived impact of the CC in improving the organization’s ability to respond in case of radar loss; b) collect proposals of improvement actions the ANSP should apply in order to manage the radar loss described in the scenario. The main results of the workshop are summarized in the following section, with focus on the after crisis scenario.

3 RESULTS

During the real event occurred in 2017, the frozen radar screens prevented all the ATCOs from visualizing the evolution of traffic for a considerable amount of time. Despite such critical situation, different sources of resilience allowed a full recovery to the normal ACC functionality in less than four hours, with limited impact on the business continuity of some Regional Airports and no negative effect on the safety of air transportation in the concerned area. For each of the thematic sections selected from the Concept Card (i.e. Adaptive capacity, Resources, Operational Margins), it was possible to provide examples of the sources of resilience that were already available in the organisation and the ones that were specifically activated to face automation loss event.

3.1 Adaptive Capacity

For what concerns the “adaptive capacity”, two different adaptation strategies were identified: (1) the reorganization of existing roles and the coordination with other Air Traffic Control units to manage the radar loss. Actually, in the first moment the automation loss, there were three Supervisors available in the operational room, who were carrying out their usual tasks. After a while, the supervisors realized that it was more efficient if they split their responsibilities in three different types of tasks. The first one started to take care of flight data management (i.e. to associate the single flights to the corresponding flight plans). The second one was enrolled in phone coordination activities with the nearby ACCs to ensure the most efficient cooperation. Finally, the third one was assisting the ATCOs at individual controller working positions. Such dynamic reallocation of tasks proved to be very effective in sustaining the effort of the whole ACC to maintain an acceptable level of continuity of the Air Traffic Control service. (2) The second adaptive capacity strategy was sustained by the coordination of the ACC with the Flow Management positions at Local, National and European level. Such coordination was necessary to modify the flight plans of all the aircraft expected to enter in the concerned Area of Responsibility (i.e. the airspace portion) of the affected ACC. All these aircraft needed to be diverted/rerouted to bypass the geographical areas affected by the radar loss and this was achieved very quickly thanks to such coordination.
3.2 Resources

The management of the loss of radar information immediately caused an imbalance between the capacity of the ACC and the demand in terms of air traffic. This required more resources to be called into duty, including both human and technical ones. For what concerns the human resources, the ATCOs involved in the management of the contingency situation were firstly helped by the ACC Operational Chief and Supervisors. Then, additional help was provided by the ATCOs that were on break during the event. According to regulations, at least the 30% of the overall ATCOs workforce needs to be available during breaks. This is a standard ‘buffer’ representing a source of resilience available by design. For what concerns the technical resources, some support was offered by the so called flight progress strips. They are small strips of paper that contain the essential information of the flights in the area of responsibility of the ACC. They represent an older type of controlling tools that were used in the past when the controller working positions were less sophisticated and did not include all the information in digital format, as in today’s multi-radar tracking systems. Nowadays, the strips are only used in combination with the fallback system, i.e. a backup system which is automatically activated in case maintenance operations or radar loss, or on ATCOs request. During a typical radar loss scenario, the fallback system has the capability to process radar data independently from the main system and to display them in less sophisticated and smaller screens. However, also the screens of the fallback system were frozen during this event, therefore only the capability of printing the flight progress strips could be used. Finally, another technical resource was used to manage the contingency, despite not being anticipated at all by official procedures. This was a simulation and training platform, located next to the main operational room, whose screen was working properly. During this situation, one of the Supervisors intensively used it to help the ACTOs in controlling the traffic at their own working position, by providing them with constant updates on the position of individual flights.

3.3 Operational Margins

In order to face with the contingency situation, different strategies were put in place to modify the normal operational margins, also through negotiation with other air traffic control entities. For the sake of brevity, we here mention only two of them. The first one was actually implemented, while the second one came out not to be necessary, due to specific circumstances on that day.

The first strategy consisted in negotiating with some of the nearby airports a different altitude for the flights that were being transferred. For example, one of the major airports in the areas was used to take under its control the descending traffic received by ACC “A” at an altitude of 90,000 feet (about 27 Km). During the contingency, it was agreed that the ATCOs working in the Control Tower of that airport would have taken the control much earlier – i.e. at 200,000 ft. (about 60 Km) - in order to reduce the workload of the ATCOs of ACC “A”. While the second strategy was the modification of the operational layout of the ACC “B” and “C”. Both centres had their radar functionalities fully operating and could provide support to ACC “A”. As mentioned before, the management of traffic is organized in sectors (i.e. specific volumes of airspace), each one controlled by a couple of ATCOs. When the traffic is very low, just a few sectors are sufficient. On the contrary, when the traffic is very high, the capacity of the ACC is increased by opening more sectors. In this case ACCs “B” and “C”
could have opened more sectors at the boundary with ACC “A”, in order to have more ATCOs providing support to the centre affected by the radar loss. At the same time, they could have gradually closed one or two sectors in the airspace volumes farther from ACC “A”, to compensate for the reallocation of their ATCOs in a different area. Such reconfigurations of the Air Traffic Control system are always possible and intrinsically make it more resilient. In that specific day, however, the second strategy was not needed as all the three ACCs had a sector configuration that exceeded the demand of traffic at that time of the day, because they were testing a new procedure.

4 DISCUSSION

The application of the principles of the DARWIN guideline and of the triggering questions included in the CC “Identifying Sources of Resilience” showed how it is possible to think about the resilience capabilities of an organization managing a critical infrastructure like the Air Traffic Control, when experiencing a major degradation of its automated systems. Interestingly enough, the identified sources of resilience concern both the reorganization of tasks and resources inside the same organization (the ACC “A”) and the reconfiguration of resources in different organizations (the neighbouring ACCs, the regional airports and the Flow Management positions at local, national and European level). Some of the resilience strategies are already embedded in the design envelope of the system (e.g. 30% of ATCOs to remain available during the break), while other strategies were identified tactically during the development of the contingency (e.g. the use of the radar screen of the simulation and training platform located close to the operational room). Overall, the personnel of the affected ACC and of the other units showed important abilities to put in place strategies compensating for the lack of the automation support, which were not fully anticipated in the official procedures. The questions still remain open: should such capabilities be incorporated in the design of the system (e.g. via improvements to the equipment and the procedures) as a response to specific events? Or rather, should these capabilities be enhanced by training, encouraging out-of-the box thinking and generation of ad-hoc solutions by the personnel? There is, of course, no univocal response, but it is very important to be aware that both strategies are needed when dealing with degraded automation scenarios affecting critical infrastructure.

REFERENCES

DARWIN. Expecting the Unexpected and Know how to Respond. Deliverable D4.3 Pilot’s Implementation and Evaluation (https://h2020darwin.eu/project-deliverables/).


Human-centred design recommendations for automated car in transition phases

Pauzié Annie¹ & Ferhat Lyess²

Ifsttar/Lescot, Cité des mobilités, Allée Général Benoist "Parc du Chêne"
Case 24 - 69675 -Bron Cedex / France

¹annie.pauzie@ifsttar.fr
²lyess.ferhat@ifsttar.fr

ABSTRACT
This study investigated drivers’ needs and preferences for the HMI design features related to the dialogue between autonomous car and user during the specific transition phase switching from automated to manual driving. Indeed, this phase is crucial in terms of road safety as the human will have to be physically fit from perceptual, cognitive and motor aspects and fully aware of the external situation, to be able to take control of the vehicle again. Several focus groups involving seniors, novices, experts, and mixed population have been conducted in order to investigate needs related to these HMI features in relation to driving experience and age. Results allow identifying preferred perceptual modalities for each group in relation to the increase emergency of the takeover. Analysis of participants’ drawings and propositions have been conducted in order to specify ergonomic recommendations for system developers.

Keywords: Human-centred design, autonomous vehicle, HMI design, focus group, seniors, road safety

1. INTRODUCTION

The concept of autonomous car is nowadays wide spreading, with a great hope in terms of road safety, replacing the human driver by reliable technology. But the presence of a human driver in the car at this stage of technical development stays necessary, as the entire road network will not be equipped, requiring transition phases between automation and manual control according to the zone. Then, as pointed out by several experts for many decades, until the driving task will be wholly automated on the full road network, there will be still an appreciable role for the human driver (Aleandri & Moyer, 1992; Barfield & Dingus, 1998; Fenton, 1970; Hancock & Parasuraman, 1992; Sheridan, 1970). In this framework, automated vehicles will create special challenges for the design of the Human-Machine-Interaction to allow safe transitions, to avoid potential confusion between human and automated system actions and to bring the driver back into the loop when necessary in a safe way (Fleisch, Schieben, Schoemig & Strauss, 2011a). Previous research on the safety of transitions from automated to manual driving has focused on assuring that the driver was aware of the transition, pointing out the importance of the design of the vehicle–driver interfaces (Fleisch et al., 2011b). Further studies have shown that drivers’ ability to take over when transition to manual control is required can be improved by providing information about the limits and the reliability of the automation (Seppeltand & Lee, 2007; Beller, Heesen, & Vollrath, 2013).

In order to overcome the potential negative impact due to non-acceptance or misunderstanding between user and automated system, HMI of autonomous vehicle will have to be intuitive, self-explanatory and well adapted to the reality of the activities conducted by the driver, especially before these critical phases of transition. Several studies showed the importance of the user-centred approach in this context of design and implementation of technology in automotive context (Pauzie & Amditis, 2010). User-centred design principles state that human specificities should be taken into consideration at each step of the system development, from the design concept till the evaluation of the prototype, but it is important to correctly understand users ‘expectations at the very first stage of the concept development. Indeed, a concept correctly designed at the early stage will require less iterative processes to improve further versions of the system and will limit the weight of further iterative improvements. So, in order to reach this goal, this study aims to involve future users at the early stages of the development of the autonomous vehicle HMI and to gather their opinions and preferences in order for the human factors team to set the ergonomic HMI design recommendations for the prototype. This approach has been recommended by the ISO norm 9241-210 “Ergonomics of human-system interaction, part 210: Human centred design for interactive systems” (2010) and demonstrated its efficiency in
HMI design processes (Pauzié, 2012). In this framework, we investigated propositions of HMI design made by samples of drivers, diversified by their age and their driving experience, related to transition phases scenarios from automated to manual driving, with two levels of situation awareness for the driver during the autonomous phase. Results will allow setting up Human-centred design recommendations for the engineers in charge of the HMI development for an autonomous vehicle prototype in the framework of a French research project named “Autoconduct” (2017-2020).

2. CONTEXT OF USE AND HMI INTRUSIVENESS REQUIREMENTS

2.1 Use case and impact on HMI intrusiveness requirements

In this study, the guidelines stated in the ISO norm 9241-210 have been followed. The first step of this recommended normalized process is to understand and to specify precisely the context of use covered by the scope of the investigation. The defined purpose of the study is to identify drivers’ needs and preferences related to HMI design, in terms of perceptual modalities and nature of the content, during the transition phase from autonomous to manual driving. This transition phase is crucial for road safety, as the driver has to be fully able to take back control of the vehicle, whatever was the driver state during the autonomous phase of the trip. The challenge of the HMI in this situation is related to its adequate level of intrusiveness, in order to attract driver’s attention, without being too much disturbing. This level of HMI intrusiveness will have to evolve in relation to the level of emergency of the situation, as the time to take control back for the driver will become more and more closer. In this framework, we identified several successive steps (figure 1): anticipation (soft informative HMI informing the driver about the happening of the transition phase), preparation (2 possible levels of HMI according to the diagnosis made by the system about the driver’s reaction to the first informative HMI: driver is in correct posture or driver is not ready), then additional steps based upon the various potential levels of driver’s readiness and responses diagnosed by the system with an increase level of HMI intrusiveness in relation to the increase of the emergency.

![Figure 1: The various steps and the corresponding HMI in the context of use in the transition phase from autonomous to manual driving](image-url)
2.2 Level of driver’s situation awareness and impact on HMI intrusiveness requirements

Studies have shown that with high automation levels, the engagement in non-driving-related tasks will increase (e.g., Jamson, Merat, Carsten, & Lai, 2013). Thus, investigation on driving takeover safety should also involve consideration on the role and the impact of non-driving-related tasks previously run by the driver. In this framework, HMI will have to be more or less intrusive based upon the nature of driver’s activities conducted before the manual takeover and their consequent impact on driver’s state and situation awareness. In order to investigate this issue, we proposed two contexts of use corresponding to two levels of driver’s situation awareness: in one case, the non-driving tasks during autonomous driving were free, going from playing game, to reading or working on the computer; in the other case, the non-driving task was defined as the driver sleeping deeply. The purpose was to make participants setting up HMI design propositions according to the level of situation awareness corresponding to each of these two situations.

Context of use specifications were very useful to set up the frame of the study, according to the possibilities of driver’s reactions and the related HMI design requirements.

3. FOCUS GROUP METHODOLOGY

It is the methodology of focus group that has been chosen to gather qualitative data about these needs and preferences regarding the HMI design. This method is an established and accepted part of the range of methodological tools available to academic researchers (Parker & Titter, 2007). It consists in gathering individuals to openly discuss a particular issue, with one or two moderators guiding the discussion and leading the group through a number of activities (Caplan, 2010, Morgan, 1998). It is a well-established technique in market research for the designing of new products (Langford & McDonagh, 2003), as well as for human factor research and usability evaluation (Jordan, 1998, McKenna, 1990, Bruseberg and McDonagh-Philp, 2001, Bruseberg & al., 2002). This workshop format appears to have several advantages in terms of methodology such as minimal participant training, collaborative group activities, individual and collective creativity and free expression.

3.1 Participants

There is a debate about the type of profile to choose to participate in focus group. Some authors recommended having homogeneous group of people (Ivanoff & Hultberg, 2006). In this case, participants share social and cultural backgrounds, they may feel more comfortable talking to each other and also are more likely to talk openly. Heterogeneous group of people have been also identified as working favorably by some other authors (Litosselti, 2003; Hennink, 2007) because it is an efficient way to maximize the possibility of exploring topics from different perspectives (Kitzinger, 1995).

We consider that each point of view has relevant aspects so we decide to set up:
- 4 homogeneous groups of people: one group of seniors (over 65 years old), one group of novice (less than 2 years of driving license), one group of expert drivers (more than 10 years practice), one group of “professionals” (drivers obliged to drive intensively in the framework of their job).
- 2 heterogeneous groups of people mixing a combination of the 4 criteria above: senior, novice, expert and professional.

For a total of 6 groups with 8 participants each, so an overall sample of 48 drivers.

Seniors:
In an ageing society, where the number of people over 65 years old is projected to double between 2010 and 2050 (Lanzieri, 2011), seniors users should imperatively be taken into account while designing innovative systems, in order to design adapted products to this population, especially in the context of road safety (Davidse R.J., 2006). Indeed, elderly drivers have specific functional abilities due to age, in terms of perception, cognition and motor control, with a high sensitivity of their level of performance in relation to time constraint (Pauzie, 2015). For example, they have more difficulties with maneuvers related to gap acceptance for crossing non-limited access highways, and high-speed lane changes on limited-access highways (Wang & Carr, 2004). They have also difficulties in attention sharing between several informative sources (Emmerson & col., 2012). So, we consider it was important to identify their needs and requirements in terms of HMI for the autonomous
vehicle, in order to design adapted product that fit with this part of the population, knowing that the seniors could benefit greatly from autonomous driving concept in the future. As a matter of fact, automation could be helpful to maintain seniors’ mobility, and consequently seniors’ health (as there is a strong relationship between the two: Dickerson et al., 2007), but this objective will be reached only if the vehicle design fits with their functional abilities.

**Novices:**
The novice drivers have difficulties in self-calibration, hazard and risk perception due to their low level of practice (Ivers et al., 2009; Konstantopoulos et al. 2010). For example, novice drivers do not show awareness about road complexity in comparison with experts, suggesting that they fail to attend to potential dangers involving the behaviour of other road users (Underwood, 2007). In the case of transition phase, when the driver will have to take back control of the vehicle, it is possible that novice drivers need more informative HMI about the road context such as traffic density and complexity of the infrastructure than the other drivers, in order to compensate their lack of experience.

**Professionals:**
We consider it was interesting to gather opinion and propositions from “professionals”, as they can have specific needs related to their intensive driving activities under job constraints. Some authors in industrial design identified the value to involve regular users, defined as “lead users” who could expect to benefit significantly from a solution to their needs, as they are good candidates for drawing the attention of designers and developers to potentially important design requirements (Herstatt and von Hippel, 1992). Indeed, taking into consideration the fact that “professionals” are driving a lot under time constraint and with potential fatigue, their opinion toward autonomous vehicle context and adequate HMI design in transition phase will be certainly of interest for developers.

**Experts:**
This group has a good practice of the driving activity, but no specific constraints like the professionals. Participants are middle-aged and in good health. Results from this group will constitute a baseline that will help us to define the potential specificities of the other groups.

### 3.2 Session

Each session was 3 hours duration. The first 2 hours were dedicated to investigate image, feeling, fears and expectations from participants related to autonomous car through various exchanges, discussions and questionnaires. The full session was video recorded, the recording serving as a material for the further analysis by the human factors team. In addition, facilitators observed the participants to make notes of the issues, the debates and the ideas that were produced,

In this paper, we are going to focus only on the final part of the session, where participants were asked to imagine, to create and to describe the design of an HMI for 2 identified use cases presented in the following section.

### 3.3 Experimental protocol

Each participant is asked to join an other one to set a pair. The 4 teams are distributed in each corner of the room and one of the 4 use cases is given to each team (among the 4, only 2 use cases will be discussed in this paper for a question of available space). Based upon the context of their use case description, participants are asked to describe and to draw the HMI design they imagine using pencils and the sketch of a car dashboard.

### 3.4 Use cases

The use case description was the trigger element to aid the participants to imagine what was the context of use and what type of HMI design would be then relevant according to their view and preferences. We did not demonstrate any solutions in advance on purpose, in order for the participants to feel free to express their creativity without any external influence.
The human factors team’s objective was to gather elements of the HMI features proposed by the participants based upon 2 use cases involving 2 different levels of driver’s situation awareness in the transition phase. The description of the use cases was handed to each pair of participants as a support for their discussion and propositions. The texts were the following:

**Use case n°1:**
“Your car is under autonomous mode for several hours and you reach the motorway exit you want to take. As planned, the car is asking you to take back control of the commands. How do you imagine this dialogue between you and your car?”

**Use case n°2:**
“Your car is under autonomous mode for several hours and you are lying on your seat, sleeping deeply. The car indicates you that the planned motorway exit is close. How do you imagine this dialogue between you and your car?”

It has to be noted that for the non-related-driving tasks in the Use case n°1, most of the participants imagined to have recreational activities, such as reading, playing Sudoku or other games, except the “professionals”, and some participants of the expert group, who did not mention these activities, as they were willing to take this free time to work. The use of the phone for personal and professional purposes was widely mentioned by all the groups, including the seniors one. Most of them mentioned also looking at the countryside, which revealed activities corresponding to a high level of situation awareness in this case, with a state facilitating the next step of manual driving takeover. When they were done, each pair of participants presented their proposition through sketch, drawing and verbal explanation to the group. Their propositions were presenting the successive steps of the visual, auditory and haptic HMI according to what they would have expected in the described situation of the use case.

*Figure 2 : Two examples of participants production concerning proposition of HMI design.*
4. RESULTS

4.1 HMI design in terms of perceptual modalities, technology and location

Propositions from participants have been analyzed and categorized according to the perceptual modalities involved, the corresponding technological and message displays and the locations in the vehicle:

- perceptual modalities such as visual, auditory, olfactory and haptic,
- corresponding technological and message displays such as screen, HUD for visual; vocal messages, music or sound signal for auditory; scent for olfactory, vibrations for haptic,
- locations in the vehicle such as car ceiling, around the windshield or on the steering wheel, on the windshield, on the dashboard or on the mirrors for visual; speakers in the car interior for auditory; car interior for olfactory and in the seat for haptic.

These schemas displayed the HMI propositions for transition phase while drivers are awake but involved in non-driving-related tasks (figure 3) and for transition phase where drivers are deeply asleep, having then a very low level of situation awareness (figure 4). Bubbles’ size is proportional to the number of propositions, giving then an insight about the importance of each HMI features based upon the participant’s point of view. In both cases, the bubbles corresponding to auditory modality are encircled in green, showing the fact that, even if the number of propositions was similar to the visual ones, the participants insisted on the fact that this modality was more important. Music is the preferred mode of auditory information for the HMI to awake drivers (figure 4), with the justification that it is less stressful that sound. Advantages of vocal messages in the two situations are clearly stated, as it is a relevant way to give details about the coming events and the recommended actions to take. In relation to the low level of situation awareness in the second case, the haptic modality with the vibration of the seat is highlighted as a modality playing an enhanced role if auditory and visual ones happened.
to be not enough efficient, in order to ensure awaking when the time to take control back becomes close. Olfactory modality such coffee aroma has been also suggested to awake driver.

4.2 HMI design propositions in relation to participants’ age and experience

All the participants’ propositions were involving both auditory and visual modalities of information for the HMI, insisting on the complementarity of the two modes. Seniors drivers insisted more than the other groups on the advantage of displaying both modalities even if the message content is redundant. The auditory mode was preferred for most of the participants, as it was considered as being the most adapted perceptual mode to attract their attention. Indeed, there was a clear understanding from them that they were not having too much availability for the visual perception because they were involved in non-driving tasks, or even when they were having no activities and daydreaming. All the drivers put forward usefulness of alerting sounds to alert them in emergency situations when actions are imminent. Only seniors insisted to have also some visual information in this context to confirm their perception, as most of them seemed to be a little afraid to have only one perceptive modality to rely upon. Their concern was to miss the auditory message or sound, as its emission is limited in time, while the visual display can stay displayed. In emergency situation, suggestions were that visual messages can present some characteristics of emergency such as the pictogram from road sign indicating “danger” with an exclamation mark in a red triangle or LED lights flashing in red, a color widely understood as referring to risk.

The visual messages were valued in their capacity to give precise and detailed information to be consulted anytime by the driver at will. Participants proposed to have this type of visual messages during the anticipation phase, allowing them to know about the time left before the takeover maneuver and the related conditions to it. The professional drivers asked to have full information about the traffic surrounding, the type of road they will have to take, and some of them refer also to information on weather conditions, revealing that this group understood the importance of situation awareness recovering after the autonomous phase to be fully operational again, due to their high driving experience.

One of the requirements of the novice drivers was to have a kind of “avatar” displayed on the central screen of the dashboard, or even on the windshield with HUD technology, in order to get a dynamic and personalized dialogue, with vocal messages and vocal recognition, during the anticipation phase. Indeed, some novice drivers, around half of the sample, happened to be not fully comfortable to give control to the autonomous vehicle, while the other half seemed to be relieved in entrusting the driving responsibility to the system. In both cases, certainly linked to personality profile, a personification of the vehicle seems to answer to their concern. In the first case, the novice driver is reassured by the “human like” dialogue of the system, while, in the second case, the novice driver has the feeling to be comforted in his choice of trusting the system because it keeps him informed about important information. This preference would be the result both of to lack of driving experience and of the generational culture for the novice group. It has to be noted the seniors expressed also the same type of concern related to trust in the system, with the split in two categories of personality such as the novice group: one happy to rely on the autonomous system, to get some rest during long trip, and the other one not feeling a high level of confidence and trust to the reliability and the safety of the automation. Without referring to “avatar” like young people, some seniors formulated the wish to have a soft personalized voice to reassure them.

Most of the participants, whatever age and experience, stressed the importance to be informed well in advance before the takeover, in order for them to get ready, especially of course in the use case with low situation awareness. They stressed the necessity to have plenty of time before the maneuver, in order to be fully prepared, to remove presbyopia glasses, to put away books and phone,... During this preparation phase, the suggested HMI would be “soft” such as music and/or blue LED light. They propose to display this type HMI to begin with and to increase little by little its intrusiveness, in order to raise their awareness progressively and smoothly. This concept of progressive HMI intrusiveness is even more highlighted in the situation where the participants were sleeping, with a high concern about setting up sweet wake up conditions.
5. CONCLUSION

Taking back control of the vehicle after a period of automated driving raised serious issues in terms of road safety related to driver’s capacity and awareness of the road context (Merat et al., 2014). During this phase, the human will be monitored by the system, with successive diagnoses to establish his/her level of awareness and ability to drive, leading to an adapted multimodal Human Machine Interaction (HMI) displays. In this framework, an important care will have to be devoted to investigate drivers’ expectation and needs to ensure there will be an efficient cooperation between the Human and the Automated System via the HMI displays, especially before and during this transition phase. In this purpose, this study allows gathering HMI propositions design of a diversified sample of drivers, varying by their experience and their age. Use cases investigated were presenting two specified levels of situation awareness for the driver, involved in non-driving-related tasks in one case, or deeply sleeping in the other case. Results indicated specific needs in relation to age and experience of driving, but also related to job constraints for professional drivers. Nevertheless, there are also requirements that are shared but most of the participants whatever these factors. The human factors team will translate these data into applicable design recommendations in order to support the engineers’ work in charge of the development of this driver’s monitoring during the transition phase.

Acknowledgment: This research has been conducted in the framework of the Autoconduct project funded by the French National Agency of Research (ANR) from 2017 to 2019.

REFERENCES


Meaningful Human Control over Automated Driving Systems

Daniël D. Heikoop, Delft University of Technology, The Netherlands, d.d.heikoop@tudelft.nl, Marjan Hagenzieker, Giulio Mecacci, Filippo Santoni de Sio, Simeon Calvert, & Bart van Arem, Delft, The Netherlands

ABSTRACT

Human Factors issues with automated driving systems (ADS) are becoming more apparent with the increasing prevalence of automated vehicles on the public roads. As automated driving demands increased performance of supervisory skills of the driver, rather than vehicle handling skills, a mismatch occurs between the demand and supply of the drivers’ skillset. Therefore, it has been suggested that drivers should at all times have meaningful human control (MHC) over ADS. The basic idea behind MHC is derived from the debate on autonomous weapon systems, and entails three essential components: human operators are (1) making informed, conscious decisions, (2) sufficiently informed about lawfulness of an action and its context, and (3) properly trained, to ensure effective control over the use of ADS. This paper presents definitions, components and potential human roles within ADS, from an interdisciplinary and a MHC perspective. The ideas presented in this paper are valuable to both designers, manufacturers, and road operators, as well as policy makers, driving licensing bodies, and lawyers and insurers, and our future research into these topics will deliver usable results for all stakeholders.

Keywords: Meaningful Human Control, Automated Driving Systems, psychology, philosophy, traffic safety.

1 INTRODUCTION

1.1 Background

The call to address Human Factors issues with automated driving systems (ADS) is becoming increasingly important, as market penetration of (partially) automated vehicles (SAE level 1 or 2; SAE, 2016) also increases. As a driver relinquishes tasks to an ADS, the driver is being ushered into a new role: that of a supervisor. During level 1 and 2 automation, longitudinal and/or lateral control is being taken over by advanced driver assistance systems (ADAS), which entails systems such as adaptive cruise control (ACC), lane keeping assist (LKA), or park assist, but also traction control and anti-lock braking system (ABS). Higher levels of automation (SAE level 3-5), will be able to take over the complete dynamic driving task, leaving the driver to merely watch the car drive itself. With this role change, the driver is subjected to out-of-the-loop difficulties (Gold et al., 2013; Louw, Merat, & Jamson, 2015). However, merely keeping the driver in the loop does not seem to be a sufficient solution, especially for prolonged periods of time (cf. Mackworth, 1948; Szalma et al., 2004). A more promising solution would be a case in which the human driver is asked to perform within their capabilities. This raises demands for meaningful human control (MHC; originated from the field of autonomous weaponry [Future of Life Institute, 2015]) over automated driving systems. The concept of MHC expresses the extent to which a human can maintain control over an automated system, even when not in (full) operational control (e.g., when an ADS rather than a driver performs operational actions; Santoni de Sio, 2016; Santoni de Sio & Van den Hoven, 2018). MHC is ultimately more demanding as well as more inclusive than the classic notion of “direct” operational control.
control, where a physical link is constantly required between human controller and controlled system. It is more demanding, as it prevents certain systems (such as vehicles) to be deemed under human control simply because somebody is ‘in-the-loop’. It is more inclusive, as it can also include supervisory control, which entails monitoring an intelligent system that is in (full or partial) operational control, and gives the user the ability to undertake action if required. Moreover, MHC applies in principle also to automated systems without direct supervision.

1.2 Objectives

This paper will address the importance of incorporating MHC in the design, implementation and use of ADS. First, we will provide a definition of MHC, after which we will introduce the core components of an ADS that incorporates MHC. Consequently, based on these core components, the chain of control within an ADS will be discussed. Lastly, the implications an ADS with MHC has on the human driver will be investigated. The implications and impacts of MHC over ADS to various types of stakeholders will be discussed, and suggestions will be made in light thereof.

2 INCORPORATING MEANINGFUL HUMAN CONTROL IN AUTOMATED DRIVING SYSTEMS

2.1 Defining Meaningful Human Control

We approach the definition of Meaningful Human Control (MHC) over automated driving systems (ADS) by distinguishing two conditions, namely tracking and tracing (Santoni de Sio & Van den Hoven, 2018). In their article, Santoni de Sio and Van den Hoven (2018) identified tracking as the first necessary condition of MHC, and defined it as the ability for a decision-making system (such as ADS) to at all times be responsive to (i.e., ‘to track’) the human agent’s (e.g., driver) relevant reasons to act. This can be at the level of a planned destination, or conventional social moral reasoning.

They identified tracing as the second necessary condition of MHC, and define it as the possibility to trace an automated system’s behaviour back to some (responsible) human agent (e.g., operator, supervisor, designer, etc.). In order for that condition to be satisfied, there has to be at least one human agent in the system’s design- and use history who is able to understand both the capabilities of the system to a sufficient extent, and their own role as targets of potential moral consequences for the system’s behaviour.

With these two definitions in mind, we aim to identify core components involved in MHC over ADS, and the roles humans could or should maintain in order to achieve MHC over ADS.

2.2 Core components of automated driving systems with Meaningful Human Control

Driver, Vehicle, Infrastructure, and Environment have been identified as categories, wherein core components that are relevant to ADS have been identified (Calvert, Heikoop, & Van Arem, submitted). Based on an extensive literature search, the underlying components enable MHC to be analysed and integrated into the ADS in a very practical fashion. These core components play an important role in the chain of control within, and interaction between the driver and the ADS.

To give a few examples, some of the core Driver components identified are the sensory components necessary
for a driver to perceive stimuli (e.g., visual, auditory, and tactile components), and components involved in the cognitive and decision making process, such as interpretation and tactics. These (dynamic) components are commonly collated as driver behaviour, as an expression of a driver’s state (as opposed to the usually static driver trait, like their personality). The (quality of) performance of these behaviours depend on whether they are automatized (i.e., skill-based), learned (i.e., rule-based), or ‘on-the-spot’ (i.e., knowledge-based) (cf. Rasmussen, 1983), and will be discussed further at section 2.3.

Some of the core Vehicle components are for instance sensors, such as the speed indicator and fuel gauge, and actuators such as the engine and clutch. Note that for an automated vehicle, for example its sensors take on a different role, from distributing information to the human driver during manual driving, to using this information itself when driving automatically, leading to completely different components for the same purpose.

For Infrastructure and Environment components, the authors identified physical and digital components, such as road surface materials and structure, and GPS and V2V communication, as well as weather components, such as rain and snow, and geographic components, such as urban or rural surroundings.

Furthermore, in their paper, Calvert, Heikoop and Van Arem (submitted) show how and why control is affected during the transition from manual to automated driving. They identified that in particular at higher levels of automation (i.e., SAE level 4-5), uncertainty increases over how this chain of control is (meaningfully) maintained, as operational control primarily lies with the ADS. Control and responsibility are shown to be even more unclear in the intermediate levels of automation (SAE level 2-3), where the chain of control is often shared between the driver and the ADS.

When aiming to adhere to the two necessary conditions of MHC for the design of ADS, following the chain of control, identified within the core components of ADS, allows explicit traceability of MHC to be performed.

2.3 Exploring human control over automated driving systems

The previous two sections explained two key conditions for MHC, and emphasized that following the chain of control within an ADS is important to apply MHC therein. This section explores the human role within ADS, and its relation towards MHC.

From a human perspective, controlling an (automated) vehicle requires skill-based, rule-based, as well as knowledge-based behaviour (Rasmussen, 1983). Loosely defined, skills are elements one is completely familiar with and can execute almost effortlessly (like steering with the wheel, negotiating a curve, or braking comfortably for a traffic light; basically anything one will be taught to do for getting your driver’s license, and now can do automatically). Rule-based behaviour entails behaviour that is performed by remembering how to act given a specific situation (such as driving 50 kph at a 50 kph road, or stopping at a red light, but also not driving through a military convoy). Knowledge-based behaviour is addressed when a driver experiences a new, unfamiliar situation, and entails drawing conclusions from past experience and general knowledge, to be able to deal with this unfamiliar situation (like driving in snowy weather, or correcting while skidding). A lot of these situations can be trained, such as during advanced driver training courses.
With automated driving, many (if not all) of these behaviours will be taken over by the ADS, meaning that there will be little (if anything) left for the driver to perform within the context of operational control. But on the other hand, drivers of such vehicles will need to learn new skills too, such as driving (interacting) with ADS, and taking on a supervisory role. However, it is as of yet unclear to what extent the decrease and/or increase in behaviours occurs during automated driving.

This is currently being investigated by looking at what effects the levels of automation, as defined by the SAE, have on human behaviour, quantifying the number of tasks added or taken over by the ADS, depending on its level of automation (i.e., level 0, 1, 2, 3, 4 or 5).

3 IMPLICATIONS AND IMPACTS OF MEANINGFUL HUMAN CONTROL FOR AUTOMATED DRIVING SYSTEMS

3.1 Tracking and tracing

This paper briefly stresses the importance of having trackability and traceability incorporated in ADS in order to achieve a meaningful form of human control over such systems. Having an ADS tracking a human’s intentions and moral standards avoids awkward situations in which the vehicle does something the human does not want, like missing an exit, or potentially in extreme cases, running over someone.

Being able to trace the status and actions of an ADS also helps the human understand the functioning of the ADS, and enables them to act appropriate, interacting as it were with the ADS, allowing for safe driving with a vehicle equipped with such a system.

Thus, when designing an ADS, adhering to the tracking and tracing principles will allow a human to have a vehicle at all times under meaningful control (i.e., not plain monitoring until something goes wrong, inevitably ending out-of-the-loop). For policy makers, this type of human control also allows for morally and legally sound ADS, as responsibility can always be tracked and traced back. Perhaps most important, the end user (i.e., driver/operator) of a vehicle equipped with ADS will face an acceptable and trustworthy form of control over their vehicle, and won’t ever feel (nor actually be) completely out-of-the-loop.

3.2 Following the chain of control

An awareness of the importance of following the chain of control opens the gateway for safe and secure design and implementation of ADS, as it shows the core components involved within an ADS, and the components affected by a transition of control due to a transition in the level of automation. This allows ADS designers and legislators tracing back responsible components much easier, as the system is transparent. Moreover, ADS designers can understand the impact their systems have on the human driver as well as designing ADS such that a human driver can comfortably and safely (i.e., meaningfully) use that system. The extent to which the chain of control has to be, and is possible to be followed, will have to be further researched. For example, from a Driver perspective, would it suffice to follow it up to a behavioural level, or should we go as deep as a neurological level? From a Vehicle perspective, do we need to know all the nuts and bolts of a given vehicle? From an
Infrastructure level, will we come to a stage where we need to investigate individual pebbles to identify what went wrong? With higher levels of automation, we might need to be able to follow the chain of control further. These questions may be particularly interesting for lawyers and insurers to be answered, and for driver licensing bodies, it would be of interest to know what future drivers need to know and be able to do in various use cases.

3.3 Control challenges and impacts for a human driver

In order to maintain meaningful human control over an ADS, understanding the effect the transition of control has on the human driver in terms of trackability and traceability could be achieved by identifying what skills, rules and knowledge is being taken over by the ADS. Unexpected or unprecedented shifts in the remaining tasks for a human driver during higher levels of automation could prove disastrous for the human driver’s ability to act as a fall-back in case of emergency. Given the core components involved with manual driving, from a human perspective, mirrored against how much influence they still have during (fully) automated driving, calls for an overhaul of the current transition of control over the various levels of automation. Moreover, the fact that a driver gets ushered into the unfamiliar, unknown role of a supervisor, entails that the demand for skilled behaviour is increasingly replaced by the demand for knowledge-based behaviour. Driving licensing bodies could step into this caveat, addressing the now unfamiliar situations, incorporating those into driver training, to allow future drivers to have the appropriate amount of skills to meaningfully control an ADS.

4 ASSESSING AND IMPLEMENTING MEANINGFUL HUMAN CONTROL IN AUTOMATED DRIVING SYSTEMS

4.1 The challenges in implementing Meaningful Human Control

Having MHC over an ADS implies full awareness (or traceability) of the systems’ status and actions, and also the presence of an ADS that knows what you want (i.e., trackability). Several, if not all, core components within such an ADS are involved when aiming to maintain MHC, and trying to address all these components in order to achieve the ability for a human driver to maintain MHC over an ADS is no easy task. From a human perspective, the role transfer comes with an unavoidable and often undesired transfer in required skill- and rule-based behaviour. A transfer of control over various levels of automation that incorporates both defined conditions implies a smooth transition for the human driver to the extent that they will not be asked to perform a task they are not qualified or capable of doing.

4.2 Future research

This paper introduces the notion of MHC over ADS by means of theoretical reasoning. Empirical assessment of the key elements discussed in this paper regarding MHC are therefore yet to be assessed. In order to be able to incorporate the notion of MHC in ADS, one could think of a human-machine interface that provides the human driver with up-to-date status and action reports, and allows the driver to actively interact with the ADS. Future research could investigate whether an ADS that overtly ‘decides’, allowing for a trackable system (in contrast to a ‘black box’), increases the likelihood of MHC over such an ADS. Furthermore, for researchers exploring this
domain, it is suggested to view the transition of control from a human perspective, rather than from a technical perspective (like the SAE levels of automation). Having an ADS incrementally take over tasks from a driver to such an extent that the driver is capable of performing their new role, could set another step closer to meaningful human control over an automated driving system. As a final, hereto related, suggestion, future research could investigate which tasks lend themselves best for ADS take-over without the risk of losing driver skills.

REFERENCES


Consumers’ Perceptions towards Autonomous and Connected Vehicles: a Focus-Group Survey on University Population

Ilias Panagiotopoulos, Doctoral Researcher, Harokopio University of Athens, Greece, ipanagio@hua.gr
George Dimitrakopoulos, Assistant Professor, Harokopio University of Athens, Greece

ABSTRACT

Autonomous and connected vehicles are an emerging technology which may prove to be the next big evolution in transportation. As of now, major vehicle manufacturing industries are developing their own prototype autonomous cars with plans to eventually release this technology to market, in full scale, up to 2030. Despite enthusiastic speculation about the potential benefits of autonomous and connected vehicles, to date little is known about the factors that will affect consumers’ acceptance or rejection of this developing technology. Gaining acceptance from end users and consumers will be critical to the widespread deployment of autonomous-connected cars. In this context the present paper describes a survey conducted by means of a questionnaire methodology distributed to undergraduate and postgraduate university students in Greece. The responses from the people participated in this survey show that performance expectancy and trust in automation could be some of the key factors influencing public attitudes towards the implementation of autonomous and connected vehicles in future transportation scheme.

Keywords: autonomous-connected vehicles, consumers’ perceptions, online survey approach, university population.

1 INTRODUCTION

Recent developments in vehicle automation technology (e.g. automatic braking, automatic cruise control, intelligent speed assistance, line keeping assistance, etc) are moving us closer to increasingly Autonomous and Connected Vehicles (ACVs). In this basis the impact of ACVs could be enormous. It could help to drastically reduce road fatalities as over 90% of the road accidents have been reported coming from human errors. Moreover, new transport services could also be developed especially when vehicles are provided with connectivity in addition to automation, e.g. traffic safety related warnings, traffic management, new possibilities for elderly people or impaired people, advanced individual comfort and convenience for drivers/users. It could also result in new business models, such as car sharing services and shared mobility which could lead to a strong decrease of vehicles on our roads [1].

All these potential societal benefits will not be achieved unless these vehicles are accepted and used by a critical mass of people; thus it will be important to understand consumers’ acceptance before the arrival of ACVs on international market. In this context it is not yet clear to what extent users accept automation technologies in vehicles and what the factors and determinants of user acceptance of automation are [2].

Various researchers have previously conducted surveys on public opinion about the perception and adoption of vehicles with autonomous-connected driving technology. In 2014, Schoettle and Sivak [3] investigated public opinion about autonomous and self-driving vehicles among 1533 respondents in the United States of America,
MEASURING DRIVER BEHAVIOUR AND THE INFLUENCE OF CONTEXT

Parallel session
Heart Rate Analysis for Human Factors: Development and Validation of an Open Source Toolkit for Noisy Naturalistic Heart Rate Data

Van Gent, P., Delft University of Technology, Netherlands, p.vangent@tudelft.nl, Farah, H., Delft University of Technology, Netherlands, van Nes, N., SWOV, Netherlands, van Arem, B., Delft University of Technology, Netherlands.

Keywords: Human factors, heart rate analysis, physiological signals, signal analysis, open source

Abstract
Heart rate data are collected often in human factors studies. Advances in open hardware platforms and off-the-shelf photoplethysmogram (PPG) sensors allow the non-intrusive collection of heart rate data at very low cost. However, the signal is not trivial to analyse, since the morphology of PPG waveforms differs from electrocardiogram (ECG) waveforms and shows different noise patterns. PPG is often preferable because it can be collected less intrusively. However, few validated open source available algorithms exist that handle PPG data well, as most of these algorithms are specifically designed for ECG data. We have developed a novel algorithm specifically for PPG data collected in noisy field- or simulator-based settings. The main aim of this paper is to present the validation of a novel algorithm on a PPG dataset collected in a recent driving simulator experiment. The dataset was manually annotated, and performance of the algorithm compared to two other popular open source available algorithms. We show that the algorithm performs well and displays superior performance on the PPG dataset. Implications and further steps are discussed.

1. Introduction
Physiological data is collected in many human factors studies. Among the physiological data, heart rate data is usually included as it is sensitive to for example changes in (driver) workload (Mehler et al., 2010), stress (Healey et al., 2005), and general driver state such as drowsiness (Danisman et al., 2010). However, capturing heart rate in the often noisy conditions of either a driving simulator or an on-road field test, and subsequently analysing the complex signals, can be difficult or costly (Brookhuis et al., 2010). Low-cost commercial devices are available, but these are generally designed for sporting contexts and not specifically for scientific research. Furthermore, the proprietary nature of the firmware and software used in these devices creates problems with data reliability and reproducibility of results.

One potential solution lies in the recent advances in wearable technology and open hardware platforms, such as Arduino\(^1\) and Raspberry Pi\(^2\). The advancement in available open hardware and software provides researchers with validated, transparent and open source data collection and processing tools. There is, however, a lack of open source available heart rate analysis algorithms that are validated, easy to use and able to handle noisy data from low-cost sensors. Implementations of heart rate analysis algorithms described in research papers are often not available, poorly documented, or require substantial technical expertise to implement properly.

Since we found the open source available heart rate analysis software unsuited for the noisy field- and simulator-based PPG data we were collecting using low-cost sensors, our aim was to develop a novel algorithm

---
\(^1\) See [http://www.arduino.cc](http://www.arduino.cc)
Heart Rate Analysis for Human Factors: Development and Validation of an Open Source Toolkit for Noisy Naturalistic Heart Rate Data

that (i) functions better on this type of data, and (ii) provides an easy-to-use way of analysing heart rate data collected in the field or in simulators. The main aim of this paper is to describe the validation of the developed algorithm on a noisy dataset collected in a driving simulator. It was designed to be resistant to typical noise patterns (e.g. motion artefacts, momentary signal loss) of participants engaged in other tasks (driving simulator, on-road car experiment, bike experiment), to be capable of handling signals from low-cost off-the-shelf sensors, and to be user friendly. It has been designed to run on both wearable devices (Arduino, Raspberry Pi), as well as on Desktop computers. For a technical overview of the algorithm, its development and its availability, please see (van Gent et al., 2018).

In the rest of the paper, we first describe basic properties of the heart rate signal as they relate to data collection and analysis. This is followed by a discussion of our methods, results and concluding remarks.

1.1 Measuring Heart Rate in Naturalistic or Simulated Settings

There are two major approaches to measuring heart rate in (naturalistic) on-road or in simulated settings, which mainly differ in the physiological properties they measure.

Electrocardiogram recordings (ECG) are collected by placing electrodes on the chest near the heart. These electrodes measure the electrical activation of the heart during each cardiac cycle. The defining feature in the ECG signal is the QRS complex (Figure 1a I-III). Advantages of the ECG signal are that it directly measures the heart’s electrical activation and that it presents a strong QRS complex presence in the resulting signal (Figure 1a). A common source of noise in ECG signals are motion artefacts resulting from sensor displacement due to participant movement. These tend to fall in the same frequency range as the QRS-complexes, which can make it difficult to filter them without deforming the QRS complex (Kirst et al., 2011). In traffic related studies, ECG recordings have been used in for example (Brouwer et al., 2015; Fallahi et al., 2016; Miyaji et al., 2008).

Photoplethysmogram (PPG) recordings offer a less invasive method of assessing the cardiac cycle. These devices employ an optical sensor to measure the discoloration of the skin as blood perfuses through the arteries and capillaries with each heartbeat. PPG is typically measured at the fingertip or through wrist bracelets. The PPG
Heart Rate Analysis for Human Factors: Development and Validation of an Open Source Toolkit for Noisy Naturalistic Heart Rate Data

The heart signal consists of a systolic peak (Figure 1b-I), a dicrotic notch (1b-II), and a secondary peak called a diastolic peak (1b-III). The secondary peak may be absent in some recordings or of very low amplitude. Advantages of the PPG method are that it is low cost, easy to set up, and non-invasive (Elgendi, 2012; Millasseau et al., 2000). Ways of obtaining the PPG signal contactless through cameras have been proposed, further reducing intrusiveness (Sun et al., 2012). However, PPG tends to display more amplitude variation over short time-intervals (Figure 1c), more variation in waveform morphology, as well as contain more noise from various sources when compared to ECG measurements. This makes analysis more difficult. In the traffic domain PPG sensors have been used by for example (Jarvis et al., 2011; van Gent et al., in press; Zhai et al., 2006).

1.2 Analysing Heart Rate Data

The heart signal is often split into heart rate (HR) and heart rate variability (HRV) measures. These measures are calculated using the distance between the detected heart beats (the RR-intervals, named because in the ECG, the largest amplitude peak is called the R-wave). The heart beats are represented by the peaks in both signals (Figure 1a, b). Despite the different underlying physiological constructs that are measured, a high correlation (median 0.97) between RR-intervals extracted from ECG and PPG signals has been reported (Selvaraj et al., 2008). This makes the PPG a valid alternative for human factors studies that require non-intrusive heart rate measurements, given that validated analysis algorithms exist.

2. Methods

The algorithm used in this study was validated using a dataset collected with a PPG sensor in a driving simulator experiment (van Gent et al., in press). The dataset contained approximately 20.7 hours of PPG recordings.

The data were split into segments of one minute each. The R-peak positions in the segments were annotated manually to serve as a basis of comparison. We compared the algorithm performance to the annotated data on four variables: detected peak position, mean of the RR-intervals for the analysed segment, beats per minute computed by the algorithm, and a common heart rate variability (HRV) measure: the standard deviation of successive differences (SDSD). To quantify the accuracy of the algorithms predictions, we used the Root Mean Squared Error (RMSE), defined as:

\[
RMSE = \sqrt{\frac{\sum(y - \hat{y})^2}{n}}
\]

Where \(y\) is the ground truth value, \(\hat{y}\) is the value predicted by the algorithm, and \(n\) the number of comparisons.

Figure 2 – Figure displaying the possible errors. These are: a.) ‘incorrectly rejected’, b.) ‘missed’, c.) ‘incorrectly accepted’. Peaks marked on a correct QRS complex but not on its R-peak maximum, are also counted as ‘incorrectly accepted’. This type of error is shown in d.). Other possible mistakes are marking an R-peak at a non-maximum position (e), or incorrectly marking a diastolic peak (f).
Heart Rate Analysis for Human Factors: Development and Validation of an Open Source Toolkit for Noisy Naturalistic Heart Rate Data

Additionally, the results of the one-minute segments were plotted and three types of errors annotated (shown in Figure 2): ‘Incorrectly rejected’ means a correct R-peak has been marked as low confidence (Figure 2a). ‘Missed’ indicates a peak is present but not marked (Figure 2b). ‘Incorrectly accepted’ indicates a peak is marked where no R-peak is considered present by the human annotator (Figure 2c). Figure 2d shows another example classified as ‘incorrectly accepted’: cases where an R-peak was marked at a non-maximum position. The HRV measures are sensitive to outliers: marking an R-peak on an anomalous position affects these measures since they rely on the intervals computed between R-peak positions. Marking a peak at an incorrect time position creates a deviation in the interval length. The algorithm has been designed to minimise the ‘incorrectly accepted’ error type.

3. Results

The used dataset represents 20.7 hours of PPG recordings split into 1,240 one-minute segments. The signals were recorded at the tip of the finger as participants were driving in a driving simulator. The sensor placement did not interfere with driving. Participants were instructed to drive as they normally would.

The data set was fully annotated. A total of 92,304 peaks were detected by the algorithm. Of these, 88,830 (96%) were correctly accepted, and 2,808 (3.04%) were correctly rejected automatically. This indicates that for 99.04% of all detections, the algorithm correctly labelled R-peaks. 666 (0.72%) peaks were incorrectly rejected. 229 (0.25%) R-peaks were incorrectly accepted. A total of 304 R-peaks were annotated as missed. Most of the incorrectly accepted peaks occur either because an R-peak was marked not at the maximum value (Figure 2, e), or because a diastolic (secondary) peak is marked as an R-peak (Figure 2, f). Future updates of the algorithm aim to further reduce these error rates. Overall, the error rates were low.

We compared the performance of our algorithm with an implementation of the Pan-Tompkins QRS algorithm (Pan et al., 1985), as well as with an open source algorithm HRVAS ECGViewer3. The latter was chosen because it is one of the first hits when searching for open source heart rate analysis software on Google, and it shows high usage statistics. It is designed for Matlab, but a standalone version is also available. The Pan Tompkins algorithm is a computationally efficient algorithm widely used in ECG analysis.

The comparison results are displayed in Table 1. They indicate that our algorithm significantly outperforms the other two open source algorithms on PPG data. The RMSE of peak position is 1.64035 (milliseconds), indicates that the standard deviation of the errors between the actual and predicted peaks was low compared to the other two algorithms. The resulting RR-intervals were also more accurate compared to the other algorithms, likely due to less missed and less incorrectly accepted beats. Differences in BPM error are not very large. Since the BPM uses the mean of all RR-intervals in a segment, it is relatively robust to a few incorrectly placed peak positions. However, effects on heart rate variability measures are large. The Standard Deviation of Successive Differences (SDSD), which is a less outlier-resistant measure for how the intervals between the heart beats vary over time, shows a large error in the other two algorithms. This shows the importance of correctly identifying R-peak positions as well as identifying incorrectly labelled peaks, as deviations risk introducing substantial error to the output measures.

---

3 See: [https://github.com/jramshur/ECG_Viewer](https://github.com/jramshur/ECG_Viewer)
Table 1 – Comparison of algorithm performance on PPG dataset (N=1,240)

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>RMSE peak location</th>
<th>RMSE RR-intervals</th>
<th>RMSE BPM</th>
<th>RMSE SDSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developed algorithm</td>
<td>1.64</td>
<td>38.87</td>
<td>4.18</td>
<td>217.41</td>
</tr>
<tr>
<td>Pan-Tomkins</td>
<td>16.52</td>
<td>171.32</td>
<td>4.76</td>
<td>364.74</td>
</tr>
<tr>
<td>HRVASE ECGViewer</td>
<td>16.43</td>
<td>272.97</td>
<td>6.71</td>
<td>1067.82</td>
</tr>
</tbody>
</table>

4. Discussion and Conclusion

In this paper we have described the validation of a novel, robust heart rate analysis algorithm developed for use in human factors studies. The motivation to develop such an algorithm was that what is available is often highly technical or expensive to implement, and because low-cost commercial measurement devices offer no suitable solution for scientific purposes.

We have evaluated the algorithm’s performance on a manually annotated PPG data set and compared the performance to two popular available algorithms. Results showed superior performance on this type of data. It must be noted that these results reflect lower performance of the algorithms only on this specific type of data: PPG data collected in the field using low-cost sensors has quite different signal and noise properties compared to ECG data, for which many available open source algorithms are designed. The evaluation does show, however, that for many human factors studies our algorithm will outperform the other available open-source methods, especially when less intrusive measurements are desired or when low-cost sensors are used.

By offering human factors researchers an openly available and validated toolkit for heart rate analysis, we aim to increase their research possibilities, as well as the reliability and reproducibility of results obtained. Future steps include increasing the accuracy and functionality of the algorithm further.

References


Heart Rate Analysis for Human Factors: Development and Validation of an Open Source Toolkit for Noisy Naturalistic Heart Rate Data

Ergonomics, 52, 95–103. https://doi.org/10.1016/j.apergo.2015.07.009


Harsh Braking by Truck Drivers: A Comparison of Thresholds and Driving Contexts Using Naturalistic Driving Data

Reinier J Jansen, SWOV, The Netherlands, reinier.jansen@swov.nl, Simone Wesseling, SWOV, The Netherlands

ABSTRACT

Frequent harsh braking is an example of risky driving behaviour by truck drivers. This study explored how threshold values on longitudinal deceleration affect the detection rate of harsh braking events across driving contexts. Naturalistic driving data from the EU project UDRIVE was used to study the behaviour of 24 Dutch truck drivers. Harsh braking events were identified through longitudinal deceleration using an initial threshold of 3.0 m/s$^2$. The maximum deceleration in each event was used to stratify the events, covering a range of threshold values found in previous studies. In total 2031 events were found. For each speed limit the mean event rate was calculated across drivers. The event rate at urban roads (30, 50 km/h) was significantly higher than at rural roads (60, 80 km/h), which in turn was significantly higher than at highways (100, 120+ km/h). Drivers with a high event rate at urban roads also showed a high event rate at rural roads and highways, but only for thresholds up to 4.0 m/s$^2$. Finally, we found distinct event rate distributions when we manipulated the threshold value. Our results suggest that driving context influences harsh braking behaviour, and that drivers have distinct driving styles. We discuss the implications for in-vehicle monitoring systems and driver coaching.

Keywords: Harsh braking, Event detection, Trucks, Naturalistic Driving, Speed limits.

1 INTRODUCTION

Heavy goods vehicles (HGV) were involved in 15% of the 25939 fatal road accidents in Europe in 2014 (Volvo Trucks, 2017). The number of deaths and injuries in traffic can be reduced by preventing risky driving behaviour. Although HGV drivers are experienced drivers that generally know how to drive safely and efficiently, they may not always use their safe driving skills to the full extent. For example, FMSCA (2016) investigated the causation of 967 accidents involving HGV in the United States. In 55% of the cases a critical reason for the accident could be attributed to the truck driver, such as failure to recognize hazards (e.g., due to driver inattention) and making wrong decisions (e.g., driving too fast, misjudging the speed of other vehicles, close following).

Safe driving behaviour can be stimulated by coaching and by giving feedback on risky driving behaviour (Horrey et al., 2012; Bell et al., 2017). Frequent harsh braking is an example of risky driving behaviour, and the focus of this study. If a truck driver is inattentive, driving too fast, or following too close, and a critical situation is imminent, then the driver will likely have to brake harshly to avoid a crash. For this reason, harsh braking events are often used to locate safety critical events in Naturalistic Driving (ND) data (e.g., Hanowski et al., 2005; Olsen et al., 2009; Zovar et al., 2014). Ideally drivers anticipate the occurrence of a critical situation, so that they do not have to brake (harshly) at all. Thus, harsh braking is a factor related to driving performance at which truck drivers can be coached. In-vehicle monitoring systems (IVMS) support coaching and giving feedback by collecting behavioural variables, such as driving speed, fuel consumption, use of cruise control, as well as acceleration and deceleration. Truck drivers can be made aware of their progress by, e.g., comparing their
Harsh Braking by Truck Drivers: A Comparison of Thresholds and Driving Contexts Using Naturalistic Driving Data

Harsh braking events are typically identified by comparing longitudinal deceleration against a threshold value. The problem is that there is no agreement on the threshold beyond which one speaks of a harsh braking event. In studies on the effect of coaching on driver behaviour, for example, Hickman & Hanowski (2011) report a threshold of 4.9 m/s$^2$, whereas Bell et al. (2017) have used a threshold of 2.3 m/s$^2$. With regard to ND studies, the NDTs project (Olson et al., 2009) uses a threshold of 1.96 m/s$^2$ to identify safety critical events, whereas the DDWS FOT project (Olson et al., 2009) uses a speed-dependent threshold of 3.4 m/s$^2$ when driving above 24 km/h and 4.9 m/s$^2$ below 24 km/h. The EuroFOT project (Malta et al., 2012) also uses a speed-dependent threshold that decreases linearly from 5.4 m/s$^2$ to 3.6 m/s$^2$ when the speed increases from 50 km/h to 150 km/h. A higher threshold will yield a lower number of harsh braking events per kilometre driven. However, little is known on whether a manipulation of the deceleration threshold yields consistent results across drivers.

Furthermore, it is likely that the driving context influences how often harsh brake events are registered. Highways and rural roads are generally more predictable than urban roads, due to their absence of pedestrians and cyclists (Wegman & Aarts, 2005). In addition, highways generally feature intersections where traffic merges in the same driving direction, whereas rural roads more often feature crossing traffic. Yet, little is known on the interaction between driving context and deceleration threshold values, either.

Our objective was to explore how deceleration threshold values affect the harsh braking event rate, and how this event rate varies as function of the driving context. Accordingly, a study was performed on the truck database of the UDRIVE project (van Nes et al., 2018).

2 METHOD

In the UDRIVE project, a fleet of trucks was equipped with multiple video cameras and sensors, through which continuous driving data were collected. We have implemented a trigger on the driving data to detect harsh braking events, which were subsequently aggregated per truck driver.

2.1 Truck drivers

Twenty-four Dutch truck drivers (23 males, 1 female) from four Dutch transport companies drove an instrumented Volvo FM distribution truck. Their age ranged from 25 to 71 years old ($M = 49.5$, $SD = 11.7$). In the period between 2015 and 2017 a total of 32831 records were collected with a travel distance between 1 and 561 km ($M = 12.6$ km, $SD = 19.6$ km, $Mdn = 7.3$ km).

2.2 Detection of harsh braking events

The truck database includes CAN data (e.g., driving speed, longitudinal acceleration, pedal use) sampled at 10 Hz, and local speed limits sampled at 1 Hz. Harsh braking events were identified with the following characteristics. First, we compared longitudinal deceleration against a threshold of at least 3.0 m/s$^2$. The resulting data segments were marked as an event if at their onset the brake pedal was depressed and the driving speed was at least 5 km/h. Multiple events within a two seconds time window were merged. For each event we recorded the deceleration peak value and the posted speed limit at the event onset. Events with
speed limits that are not part of the Dutch speed limit system (e.g., 90 km/h) were excluded from subsequent analysis, as were events of which no speed limit data were available.

2.3 Data analysis

For each driver and at each speed limit we calculated the event rate as the number of harsh braking events divided by the distance driven at the corresponding speed limit. A relatively liberal threshold value for longitudinal deceleration was selected on purpose with the aim to yield a large initial set of events. The registered peak value of deceleration was then used to group the events into five categories, based on the following cut-off values: 3.0 m/s², 3.5 m/s², 4.0 m/s², 4.5 m/s², and 5.0 m/s². This approach allowed us to examine the event rate at distinct thresholds for longitudinal deceleration.

3 RESULTS

A total of 2031 harsh braking events were identified over a distance of 227000 km. We first examine the effect of speed limits on event rate, followed by differences across drivers.

3.1 Harsh braking events across speed limits

Figure 1 displays the mean harsh braking event rate as a function of speed limit. Each stack within a bar represents the event frequency in the range set by two subsequent cut-off values. Most events were identified on 50 km/h roads, followed by 70 and 80 km/h roads. However, 30 km/h roads yielded the highest event rate, because the distance covered was small. Except for 60 km/h roads, the event rates decrease as higher cut-off values are chosen.

Looking at the total bar height, there appear to be three clusters of speed limits with a similar event rate at a cut-off value of 3.0 m/s². The highest event rates are found in urban areas (speed limits: 30, 50, 70 km/h). In rural areas (speed limits: 60, 80 km/h) the event rate is approximately 2-3 times lower than in urban areas, and at highways (speed limits: 100, 120+ km/h) the event rate is about one tenth that of rural areas (note: trucks were restricted at a driving speed of 85 km/h). At rural roads and highways the distribution of event rate across
drivers was significantly different from the normal distribution. Therefore, a Friedman ANOVA was performed, which yielded a significant effect on speed limit cluster, $\chi^2(2) = 36.55$, $p < .001$. Two Wilcoxon signed ranks tests were used for post-hoc comparisons. A Bonferroni correction was applied, such that the significance was tested against an alpha of .025. The event rate at urban roads ($Mdn = 1.92$ events/100km) proved to be significantly higher than at rural roads ($Mdn = 0.48$ events/100km), $T = 1$, $p < .001$. Likewise, the event rate at rural roads was significantly higher than at highways ($Mdn = 0.018$ events/100km), $T = 0$, $p < .001$.

3.2 Harsh braking events across drivers

We examined differences between drivers within speed limit categories, and across speed limit categories. Figure 2 shows the event rate across drivers for urban roads. The drivers were ordered based on their event rate at the lowest cut-off value (i.e., 3.0 m/s$^2$). If a higher cut-off value had been chosen, this order would have changed drastically. Similar patterns were found at rural roads and highways (the corresponding figures are omitted due to limited space). To summarize, drivers differ in how often and how harshly they brake, and harsh braking intensity appears to be unrelated to harsh braking frequency.

Figure 2 - Harsh braking event rate across drivers within urban areas (speed limits: 30, 50, 70 km/h).

To compare between drivers across speed limit clusters, we have calculated the correlation between event rates at each cut-off value. Significant positive correlations were found between each speed limit cluster at the lowest cut-off value, see Table 1. When the cut-off value is increased, however, the magnitude of the correlations and their significance declines. This finding suggests that drivers differ in where they perform very harsh braking manoeuvres, at least when the driving context is operationalized in terms of urban, rural and highway speed limit clusters.

<table>
<thead>
<tr>
<th>Location</th>
<th>Cut-off ≥ 3.0</th>
<th>Cut-off ≥ 3.5</th>
<th>Cut-off ≥ 4.0</th>
<th>Cut-off ≥ 4.5</th>
<th>Cut-off ≥ 5.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>U   R   H</td>
<td>U   R   H</td>
<td>U   R   H</td>
<td>U   R   H</td>
<td>U   R   H</td>
</tr>
<tr>
<td>Urban</td>
<td>1   .68** .67**</td>
<td>1   .51* .58**</td>
<td>1   -.15 .46*</td>
<td>1   .25 .38</td>
<td>1   -.09 .18</td>
</tr>
<tr>
<td>Rural</td>
<td>.   1   .73**</td>
<td>.   1   .39</td>
<td>.   1   -.10</td>
<td>.   1   -.15</td>
<td>.   1   -.13</td>
</tr>
<tr>
<td>Highway</td>
<td>.   .   1</td>
<td>.   .   1</td>
<td>.   .   1</td>
<td>.   .   1</td>
<td>.   .   1</td>
</tr>
</tbody>
</table>

NOTE: U = Urban, R = Rural, H = Highway. * $p < .05$, ** $p < .01$. 

Table 1 - Pearson correlation on event rate between driving contexts, stratified across cut-off values (m/s$^2$).
4 DISCUSSION

The objective of this study was to explore the impact of deceleration threshold values on the detection of harsh braking events across driving contexts. Most truck drivers in the UDRIVE database have been found to perform harsh braking manoeuvres, yet the event frequency varies across drivers. Some drivers show many harsh braking events, but the magnitude of deceleration in each event is modest. Other drivers perform relatively few harsh braking manoeuvres, but for those drivers the magnitude of deceleration is much larger. Consequently, when drivers are ordered according to their harsh braking event rate, the ordering changes when the threshold is shifted from a liberal to a more conservative value. The implication of this finding with regard to driver coaching is that the interpretation of individual driver performance compared to fleet performance depends on the threshold that is chosen to identify harsh braking events.

With regard to driving context, the momentary speed limit significantly influenced the frequency of harsh braking events. At urban roads (speed limits: 30, 50, and 70 km/h) the event rate was approximately twice as high compared to rural roads (speed limits: 60 and 80 km/h). In turn, the event rate at rural roads was approximately ten times higher than events found at highways (speed limits: 100, 120, and 130 km/h). Seeing that some drivers might drive more in urban areas and others more on the highway, a comparison between individual driver and fleet performance should be corrected for the driving context in which harsh braking events are collected. Alternatively, harsh braking behaviour could be evaluated separately for each context.

Our study examined driving behaviour by Dutch truck drivers, which may limit the generalizability of our findings towards other countries. For example, the event rate at highways ($Mdn = 0.018$ events/100km) reported in this study is in line with the event rates reported in the US study of Hickman and Hanowski (2011) for a long-haul carrier ($M = 0.0123$ events/100km) and a short-haul carrier ($M = 0.025$ events/100km). For urban and rural roads, however, our event rate was two to three orders of magnitude larger. This difference could be explained if the data in the US study were mainly collected on highways, including the short-haul carrier, but such information was not reported. An alternative explanation could be that traffic on urban roads differs between US and Dutch cities. US residents typically commute by car, whereas relatively many Dutch residents use their bicycle. Consequently, the number of interactions between truck drivers and cyclists is likely higher in Dutch urban areas than in US urban areas, which in turn may account for an increased harsh braking event rate.

Another limitation of our study is that most records in the UDRIVE database covered a distance smaller than 10 km, which is a fraction of trips driven by long-haul trucks. Long, monotonous trips increase fatigue and reaction time (Ting et al., 2008), which may increase the number of harsh braking events. However, all trips in the present study yielded maximally one event, including trips covering a large distance. Furthermore, the event rate has been stratified across speed limits, and expressed as proportion of the distance driven. Therefore, a potential bias introduced by trip distance has been mitigated to the best possible extent.

Previous studies on harsh braking behaviour report a wide range of deceleration threshold values. Some studies used a low value to find a large number of events (i.e., high sensitivity), followed by a manual validation of actual harsh braking events (Bell et al., 2017). This approach may be too time-consuming for use in automated commercial in-vehicle monitoring systems. Other studies have used a relatively high threshold value to decrease
the number of false positives (i.e., high specificity), which may be more attractive for automated processing, yet it comes at the risk of missed events. The optimal balance between sensitivity and specificity remains a subject for future research. Until then, our study serves as a reminder that an arbitrarily chosen threshold will likely influence feedback on the performance of individual drivers in relation to their peers.

ACKNOWLEDGMENTS

This work is part of the research project MeBeSafe, which has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 723430.

REFERENCES


“A Normal Driving Based Deceleration Behaviour Study Towards Autonomous Vehicles”

Stavroula Panagiota Deligianni, School of Architecture, Building & Civil Engineering, Loughborough University, U.K., Professor Mohammed Quddus, School of Architecture, Building & Civil Engineering, Loughborough University, U.K., Professor Andrew Morris, Loughborough Design School, Loughborough University, U.K., Dr Aaron Anvuur, School of Architecture, Building & Civil Engineering, Loughborough University, U.K.

ABSTRACT

Vehicle automation has recently attracted significant interest from the research community worldwide. Notwithstanding the remarkable development in autonomous vehicles (AVs), there is still a concern about the user’s comfort since most research has mainly focused on the safety aspect. To be comfortable for different users, AV should adapt its driving style to mimic the human’s one. One of the most critical factors affecting the comfort level is the braking. It is however unclear which factors affect the braking behaviour and which braking profiles make the occupants feel safe and comfortable. This work therefore aims to thoroughly explore the deceleration behaviour of drivers using naturalistic driving study (NDS) data from two Field Operational Tests (FOT), the Pan-European TeleFOT (Field Operational Tests of Aftermarket and Nomadic Devices in Vehicles) project and the FOT conducted by Loughborough University and Nissan Ltd. A total of about 28 million observations were examined and almost 3,000 deceleration events from 37 different drivers and 174 different trips were identified and analysed. With the aid of a cluster analysis, a number of homogeneous scenarios based on human factors were formed. The scenarios have led to the application of multilevel mixed effect linear models to each cluster examining all influencing factors of the braking behaviour. The results indicate a dependence of the deceleration behaviour differing due to driver characteristics, initial speed and the reason for braking. Findings from this study will support vehicle manufacturers to ensure comfortable and safe braking operations of AVs.

Keywords: deceleration behaviour, normal driving, human factors, (semi)autonomous vehicles

1 INTRODUCTION

Vehicle automation offering safer, faster and cleaner transport and trying to eliminate the human error is of great interest to the society. To achieve high user acceptance and market penetration in the domain of autonomous driving, the design of automated driving functions is crucial and should offer flexibility and adaptability (Griesche et al. 2016). One design approach for those functions is the analysis of the human behaviour and a successive implementation of the results into the autonomous systems (Deligianni et al. 2017). As pointed out in several studies, occupants do not feel comfortable inside the AVs due to the unnatural driving performance of the current technology (Kuderer et al. 2015, Elbanhawi et al. 2015, Scherer et al. 2015). Therefore, AVs should resemble human driving style, taking into consideration the individual driver’s preferences, who in AVs becomes a passenger (Elbanhawi et al. 2015, Scherer et al. 2015). Ride comfort is a subjective concept understood as a state achieved by the removal or absence of uneasiness and distress and may vary considerably among drivers, since human drivers adopt different driving styles based on the personality, the age, the gender, etc. (Kuderer et al. 2015). Additionally, a single subjective evaluation of ride comfort and investigation of ergonomics factors are no
longer considered an acceptable and competitive way to assess the passenger experience (Elbanhawi et al. 2015). Comfort is influenced by multiple factors such as temperature, vibration, time headway, time-to-collision, longitudinal and lateral acceleration/deceleration and jerk (Le Vine et al. 2015, Elbanhawi et al. 2015). One of the most important and critical factors is the braking as a sharp deceleration is closely connected to accidents.

Elbanhawi et al. (2015) reviewed the traditional comfort measures and proposed autonomous passenger’s comfort factors, e.g. naturality, apparent safety and motion sickness. Further, the gap in path planning from a passenger comfort perspective is highlighted. A study of Griesche et al. (2016) concluded that a preference among most drivers is for an AV to imitate their own driving style or a similar one. Aiming towards an increase of passenger’s comfort too, Dovgan et al. (2012) developed a multi-objective algorithm to optimise the control action with three objectives, i.e. travelling time, fuel consumption and comfort. Scherer et al. (2015) examined essential driving parameters towards the increase of comfort feelings for passengers inside an autonomous car, resulting in the parameters of the longitudinal control and, more specifically, braking and acceleration.

Several studies have investigated the factors related to the braking behaviour (Haas et al. 2004, Loeb et al. 2015, Deligianni et al. 2017). To determine the differences in emergency braking performance between novice teen drivers and experienced adult drivers, Loeb et al. (2015) conducted a simulator study, resulting in poor response and quality of braking from novice drivers compared to experienced drivers. The study conducted by Deligianni et al. (2017) examined the deceleration events from different drivers and concluded that the deceleration is mostly affected by both, kinematics factors and the reason for braking. The purpose of the study conducted by Haas et al. (2004) was to evaluate driver deceleration and acceleration behaviour at stop sign–controlled intersections. The results indicate a wide variability in rates of acceleration and deceleration and a strong relationship between the initial speed and both deceleration and acceleration.

The research to date on autonomous vehicles tends to focus on the safety aspect rather than the comfort of the passengers. Previous studies of deceleration behaviour have examined factors related either to the driver or the vehicle, but there is a lack of studies examining situational factors or considering all the factors at once (multilevel analysis) that play an important role for the driver’s decisions. As a result, the impact of these factors (driver, kinematics, situational) on the deceleration behaviour has not been fully understood yet. This study aims to fill in this knowledge gap by analysing drivers’ braking behaviour obtained from normal driving, which feels comfortable, using NDS data in different scenarios (i.e. different road infrastructure and different road conditions). It will focus on discovering the relationship between the braking behaviour and its influencing factors. In addition, a library of deceleration functions will be developed, suitable for different road conditions and road users, which can be of great use in the development of comfortable and safe autonomous braking.

2 DATA DESCRIPTION

Since the comfort of drivers is the key in this work, the data that were used reflects driver’s normal braking and does not include any safety critical events (emergency braking). The data was obtained from two different ethical approved projects: the TeleFOT project and a cooperation project between Nissan Motor Company Ltd and Loughborough University. Both consist of Field Operational Tests providing NDS data. The equipment comprises a GPS and an accelerometer linked and synchronised to a four-channel video system monitoring the drivers’
behaviour using the Race Technology Ltd with a sampling frequency of 100 Hz. The sample was composed of 37 drivers aged 18-65 with at least 2 years driving experience and the summary of the participants’ information for both projects is presented in Table 1. The participants at Nissan project were asked to drive along three specific routes (trip duration 15-25 minutes) that represented different road types, i.e. rural, motorway and urban and used three different cars (i.e. a Nissan Qashqai, a Peugeot and a Ford). The participants of the TeleFOT project were asked to drive along a specific 16.5 km long route with mixed road types in the Leicestershire area of England using one instrumented vehicle (Figure 1). All participants drove for a couple of hours to familiarize with the car. Therefore, the influence of the car type should be considered and included in the model. As a result, a total of about 28 million observations were examined from 37 different drivers, 4 different cars and 174 different trips.

Table 1 – Drivers’ demographic characteristics

<table>
<thead>
<tr>
<th>Gender/Age</th>
<th>17-30</th>
<th>31-40</th>
<th>41-50</th>
<th>51-60</th>
<th>61+</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>Female</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td>4</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>2</td>
<td>37</td>
</tr>
</tbody>
</table>

To extract the data of our interest from these projects, the Race Technology V8.5 software was used. The obtained variables were time, longitudinal acceleration, car speed, travelling distance, video frame and GPS coordinates for each deceleration event. The other necessary data (i.e. the trip duration, the maximum and the mean deceleration of the car during the event, the mean and the initial speed of the car during the event, the duration, the travel distance) as well as the detection of the deceleration events were calculated by an algorithm developed in MATLAB. By analysing the video at the moment of the events, two important variables were identified: traffic density and situational factors, i.e. whether there is a traffic light, a roundabout, a junction, a pedestrian crossing, an obstacle or a combination of the abovementioned which compose the reason of braking and whether the car stops in car blocks. The event starts when the acceleration is less than -0.5 m/s² and ends with the release of the brake. Various descriptive statistics were generated to understand these variables, the relationships between them and the effect they have on deceleration behaviour. The average maximum deceleration was found to be equal to -2.53 m/s², the absolute maximum value was 7.08 m/s² and it was observed that the car, the reason for braking, the speed and the combination of age and gender seemed to affect the deceleration value.
3 METHODOLOGY

To achieve the identified objectives, the following methodology was applied. Firstly, the deceleration events were identified from the datasets using adequate thresholds. Most of the deceleration rates observed in both projects are relatively low due to the nature of the Field Operational Test (FOT) which reflects driver’s normal braking with no safety-critical events. Therefore, the threshold was set at $2\text{m/s}^2$, which is the lowest value found in the literature to detect deceleration events (Deligianni et al. 2017). For this purpose, an algorithm was developed in MATLAB that recognises the deceleration events, divides them into two parts, i.e. the press and the release of the brake and estimates the best function out of three different typical braking patterns for both parts: (1) the driver brakes gradually, (2) the driver brakes smoothly and after presses the brake harder and (3) the driver brakes firmly at the beginning followed by a gradually smoother braking. Similar patterns were used for the break release.

The next step is the creation of different scenarios based on human factors, to reflect the differences among the drivers, and on the braking pattern. To accomplish that, a cluster analysis was employed. Traditional clustering methods, i.e. hierarchical clustering and K-mean clustering were eliminated as suitable solutions due to the data’s nature. Specifically, hierarchical clustering cannot handle a large number of observations and K-means clustering is not appropriate for categorical variables. Therefore, the 2-step cluster analysis in SPSS was used, since this method can handle categorical variables (such as gender, age categories and braking profiles) as well as big datasets. Having the deceleration events clustered and with the aim of examining all the influencing factors of the braking behaviour, the multilevel mixed effect model was applied to each cluster using the StataMP 13 software. This model is the most appropriate as the data presents a hierarchical structure (i.e. each driver conducted several trips and each trip included many deceleration events) (Figure 2). The factors that were considered are: (1) event-level factors, such as situational factors (reason of braking, traffic density), kinematic factors at the beginning of braking, etc. (2) trip level factors, such as trip duration, trip distance, the model of the car and (3) driver level factors.

4 RESULTS AND DISCUSSION

The results of the estimation of the best fitted braking pattern showed that 636 out of 2715 deceleration events followed the (1) braking pattern (gradually braking), 1104 followed the (2) braking pattern (smooth followed by harder braking) and 999 followed the last one (3) (hard following by smoother braking). The influence of the driver’s characteristics on the braking behaviour was analysed using the cluster analysis discussed in the previous section. Five clusters were created as an outcome and their features can be seen in Figure 3. It can be concluded that old people (cluster 1 and 3) slightly prefer the braking pattern (2) whereas young people also use the third

![Figure 2 - The hierarchical structure of the data](image-url)
braking pattern (3). Moreover, the different clusters present different deceleration characteristics. This can be supported by the results of the Analysis of Variance (ANOVA) test (p=0.045<0.05), conducted to test the differences between the means of the maximum deceleration for each cluster. Additionally, it was concluded from the Tukey’s HSD test that old females brake the hardest whereas old males brake the softest.

After clustering the observations into 5 groups, the maximum deceleration value was analysed using statistical analysis for each cluster. Since the driver effect has been included in the clustering, the model that was used was the 2-level linear regression model based on the trip level. The explanatory variables, which include distance, initial speed, if the car should stop, traffic density and the reason for braking, were kept the same among the clusters. The results from the analysis are presented in Table 2. The overall intra-class correlation (ICC) varies from 0.037 (cluster 1) to 0.16 (cluster 4) indicating that 3.7% and 16% of the variation in the deceleration value is explained by the trip-level hierarchical data structure. Therefore, all models show a reasonable goodness-of-fit.

<table>
<thead>
<tr>
<th>Deceleration</th>
<th>Cluster1</th>
<th>Cluster2</th>
<th>Cluster3</th>
<th>Cluster4</th>
<th>Cluster5</th>
</tr>
</thead>
<tbody>
<tr>
<td>distance</td>
<td>-0.018</td>
<td>-0.020</td>
<td>-0.024</td>
<td>-0.012</td>
<td>-0.033</td>
</tr>
<tr>
<td>initial_speed</td>
<td>-0.012</td>
<td>-0.020</td>
<td>-0.024</td>
<td>-0.012</td>
<td>-0.033</td>
</tr>
<tr>
<td>traffic_light</td>
<td>0.164</td>
<td>0.106</td>
<td>0.071</td>
<td>0.164</td>
<td>0.178</td>
</tr>
<tr>
<td>roundabout</td>
<td>0.112</td>
<td>0.133</td>
<td>0.107</td>
<td>0.164</td>
<td>0.178</td>
</tr>
<tr>
<td>t_junction</td>
<td>0.145</td>
<td>0.133</td>
<td>0.103</td>
<td>0.164</td>
<td>0.178</td>
</tr>
<tr>
<td>cross_junction</td>
<td>-0.198</td>
<td>-0.100</td>
<td>-0.217</td>
<td>-0.174</td>
<td>-0.263</td>
</tr>
<tr>
<td>other</td>
<td>0.117</td>
<td>0.017</td>
<td>0.097</td>
<td>0.117</td>
<td>0.091</td>
</tr>
<tr>
<td>stop_car</td>
<td>-0.198</td>
<td>-0.100</td>
<td>-0.217</td>
<td>-0.174</td>
<td>-0.263</td>
</tr>
<tr>
<td>_cons</td>
<td>0.037</td>
<td>0.055</td>
<td>0.07</td>
<td>0.10</td>
<td>0.11</td>
</tr>
</tbody>
</table>

The most statistically significant variables affecting the deceleration value for almost all the models are the initial speed and if the car should stop. Increasing the initial speed by 1m/s leads to a harder braking (the decrease varies from 0.012 to 0.033m/s²). Another important factor is the cause of braking. Specifically, approaching a roundabout or a junction results in softer braking compared to a dynamic obstacle, whereas approaching a pedestrian crossing leads to harder braking. Furthermore, for cluster 4 the existence of a traffic light made the
braking softer. The traffic density was revealed to be insignificant for all the clusters.

Summarising this research revealed the factors that significantly affect the braking for each cluster, showing the braking preferences for each group. Also, it was concluded that drivers can accept a harder braking because of a dynamic obstacle or a pedestrian crossing. In addition, it demonstrated that driver deceleration cannot be effectively modelled by applying average rates and one braking profile, since it varies a lot depending on the driver and the situational factors, supporting more the idea of a personalised autonomous vehicle. Haas et al. (2004) have concluded in similar results, taking though into account only the gender and the speed. The limitation of this work lies in the fact that while 37 drivers are sufficient to conduct a statistical analysis, more drivers are needed to generalise the cluster analysis. Also, even if the assumption that the drivers feel comfortable when the AV mimics their own driving style is supported by different studies (Kuderer et al. 2015, Elbanhawi et al. 2015, Scherer et al. 2015, Griesche et al. 2016), it needs more investigation. This paper concentrates on studying in depth the deceleration events in normal driving in order to support vehicle manufacturers to ensure comfortable braking operations and specifically when an AV detects a hazard and calculates the distance that it should stop at, it will be able to brake in a safe and comfortable way, considering who the passenger is and different situational factors. This will lead to autonomous vehicle’s wide acceptance and market penetration.

5 REFERENCES


Investigation of Herringbone pattern and Optical Circles for Safe Driving Behaviour at Curves Using Driving Simulator

Hammad Hussain Awan, Hasselt University, Transportation Research Institute (IMOB), Agoralaan, 3590 Diepenbeek, Belgium, hammadhussain.awan@uhasselt.be, Arne Houben, Hasselt University, Faculty of Engineering Technology, Agoralaan, 3590 Diepenbeek, Belgium, Sander Westhof, Hasselt University, Faculty of Engineering Technology, Agoralaan, 3590 Diepenbeek, Belgium, Ali Pirdavani, Hasselt University, Faculty of Engineering Technology, Agoralaan, 3590 Diepenbeek, Belgium, Muhammad Adnan, Hasselt University, Transportation Research Institute (IMOB), Agoralaan, 3590 Diepenbeek, Belgium, Tom Brijs, Hasselt University, Transportation Research Institute (IMOB), Agoralaan, 3590 Diepenbeek, Belgium.

ABSTRACT
Horizontal curves have 1.5 to 4 times higher probability of accidental occurrence than tangent sections. Majority of these accidents are caused due to human error. Therefore, human behaviour at the curve needs to be corrected. In this study, two different road marking treatments, 1) optical circles and 2) herringbone pattern, were used to influence drivers’ behaviour while entering the curve on a two-lane rural road section. The road section is selected from the Belgian Road Network and a driving simulator is used to perform experiments. Both treatments were found to reduce speed before entering the curve. However, speed reduction was more gradual when optical circles were used. Herringbone pattern had more influence on lateral position than optical circles by forcing drivers to maintain a safe distance with the on-coming traffic on the adjacent lane. The study concluded that among other low-cost speed reducing methods, optical circles is an effective tool to reduce speed and increase drivers’ attention. Moreover, Herringbone pattern can be used to reduce accidents in the curve where the main cause of the accident is the lateral position.

Keywords: Driving Simulator, Driving Behaviour at curves, optical circles, Herringbone pattern.

1 INTRODUCTION
Safety is a very important aspect of road design to be considered, especially in case of rural roads. On rural roads, certain behaviour is expected from the driver which is communicated through various clues. Knowledge of drivers’ perception of these clues is important as failure to comprehend these will result in unsafe situations. A road design can be considered as self-explaining when it is able to evoke the required behaviour from the drivers without the help of road signs (Theeuwes and Godthelp 1995). With few additional road markings, a road can be made self-explanatory.

Previous research shows that probability of occurrence of a fatal accident in curves is 1.5 to 4 times higher than that for tangent sections (Alexei et al. 2005) which makes safety a major concern in designing horizontal curves especially in rural areas. Radius of a curve is directly proportional to the design speed of the road (AASHTO 2011) and in some situations, it is require to be increased for enhancing road safety. Solutions other than changes in geometric design are required if geometry of the curves cannot be modified due to factors such as lack of available space etc. This is also the case with the two locations selected from two-lane rural highways in the Belgian road network.

2 LITERATURE REVIEW
Several pavement markings have been studied previously to make roads self-explaining at different sections of the road. To ensure safe driving through the dangerous section, speed reduction before entering the danger zone and maintaining the appropriate lane position is important. Charlton (2007) used various combinations of pavement markings and warning signs in a driving simulator study and found that herringbone pattern used with signboards increased the separation gap between the two opposing lanes of traffic and influences driver to follow the path that provides maximum available radius through the curve, which implies appropriate lateral position. Ariën et al. (2012) studied transverse rumble strips and herringbone pattern in a driving simulator and found that transverse rumble strips were more effective than herringbone pattern in reducing speed. Kerman et al. (1982) proposed a reduction in approach speed to reduce speed at curves. This is because speed choice at curves is highly dependent on approach speed and geometry of the curve.
Investigation of Herringbone pattern and Optical Circles for Safe Driving Behaviour at Curves Using Driving Simulator

Some configurations of pavement markings are presumed to manipulate speed perception of the drivers by creating an illusion of high speed (called perceptual pavement markings). Kitamura and Yotsutsuji (2015) studied effects of sequential transverse and lateral markings on perceived speed on a single-lane straight road on drivers’ behaviour using driving simulator. Different configurations of transverse markings along with roadside poles were created in which spacing between transverse markings and poles was decreasing gradually. Results indicated that perceived speed was higher than actual vehicle speeds. In this study, optical circles (created on the same principle as of optical bars) and herringbone pattern (used by Charlton (2007)) were applied on the two horizontal curves selected from Belgian road network using the similar methodology as explain in Ariën et al. (2017).

3 METHODOLOGY

The driving simulator at the Hasselt University, Transportation Research Institute (IMOB) is a fixed base medium fidelity simulator consisting of a mock up car (Ford Mondeo) with a seamless, curved screen placed at front of the vehicle. A synchronized image of 4200 by 1050 pixels quality is presented by three projectors at 60Hz refresh rate with 180° wide vision. Data from the driving simulator was collected at the frame rate. Two horizontal curves (named as Hoogstraat and Masseik in this paper) selected from Belgian road network on a two-way rural road were created in STISIM Drive Version 3. Lane width on the Hoogstraat and Masseik was 3.2m and 2.8m respectively. Both of these were transitional curves and their lengths and radii are given in Table 1. Pavement markings i.e. optical circles and herringbone pattern were applied on both of these curves. Effects of these markings were studied by comparing both curves with a control scenario in which no treatment was applied. As a result, six road sections (three sections for each curve) were created. These road sections were arranged in a randomized order to make two 18km long scenarios. The entire driving duration for both test scenarios was approximately 30 minutes.

Table 1: Curve lengths and their radii

<table>
<thead>
<tr>
<th>Sr. No</th>
<th>Curve Radius (m)</th>
<th>Curve Length(m)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hoogstraat</td>
<td>Masseik</td>
</tr>
<tr>
<td>1</td>
<td>170</td>
<td>169</td>
</tr>
<tr>
<td>2</td>
<td>94</td>
<td>92</td>
</tr>
<tr>
<td>3</td>
<td>161</td>
<td>97</td>
</tr>
<tr>
<td>4</td>
<td>219</td>
<td>688</td>
</tr>
</tbody>
</table>

Optical circles segment was 90 meters long with a centre-to-centre distance of 10 meters between circles and was placed approximately 100 meters before the start of the curve. The diameter of circles increased gradually from 1.4m to 2.3m with an increment of 0.1m (Figure 1(a)). The illusion of increased speed is created by the concept of forced perspective illusion (Endler et al. 2010). Optical circles in this study are designed on the similar principal of previous studies (Godley et al. 2000, Galante et al. 2010, Montella et al. 2011). Reason to choose circles over square and eclipse is that circles require less area than squares and eclipse (Hussain 2017).

Figure 1: (a) Top view of optical circles, (b) Top view of Herringbone Pattern,
Investigation of Herringbone pattern and Optical Circles for Safe Driving Behaviour at Curves Using Driving Simulator

Herringbone pattern used by Charlton (2007) was modified according to the road width of the roads considered in this study. Width for the drivers to drive on both road sections was kept 2.5 meters at the start of the curve. This width gradually increases to the maximum lane width in the middle of the curve and then starts to reduce again. The inclination of herringbone strips was kept along the direction of the travel. Top view of herringbone pattern is shown in Figure 1 (b). The experimentation in this study was approved by the Ethical Committee of the University of Hasselt. 49 participants (14 female, 35 male) volunteered in this study with age range between 19-54 years with mean age of 26.08 years. After detecting outliers, data for 43 participants were considered in the analysis. Driving behaviour parameters considered in this study are longitudinal speed, mean acceleration/deceleration and mean lateral position. Effects of pavement markings (optical circles and herringbones) are computed and compared for both curves on 11 points (along the longitudinal axis) selected for the analysis. Description of these points is provided in Figure 2. For lateral position values obtained from the driving simulator, the central median was considered as benchmark. Positive values indicate that driver is on the right side of the median.

4 RESULTS

Due to the difference in lane width of the two roads, both curves are analysed individually by applying MANOVA statistical test to study overall effects of independent variables (road marking, points, two-way interaction between road marking and points) on dependent variables (speed, acceleration and lateral position) and repeated measures ANOVA to study the with-in subject effect of independent variables on each dependent variable individually. Table 2 and Table 3 present the analysis for Hoogstraat and Masseik respectively. Road markings and points were found to have overall significant effect including the two – way interaction between them (Wilks’ Lambda p < 0.05). Effects of markings are explained on all three dependent variables in this section.

Table 2 Statistical analysis results for the curve Hoogstraat

<table>
<thead>
<tr>
<th>Independent factor</th>
<th>F value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road Markings</td>
<td>206.048</td>
<td>.000</td>
</tr>
<tr>
<td>Points</td>
<td>6.636</td>
<td>.000</td>
</tr>
<tr>
<td>Road Marking * Points</td>
<td>2.077</td>
<td>.000</td>
</tr>
</tbody>
</table>

Test of With-in Subject Effects (Greenhouse-Geisser)

<table>
<thead>
<tr>
<th>Independent Factor</th>
<th>Speed</th>
<th>Acceleration/Deceleration</th>
<th>Lateral Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road Markings</td>
<td>.185</td>
<td>6.342</td>
<td>.634</td>
</tr>
<tr>
<td>Points</td>
<td>112.909</td>
<td>91.303</td>
<td>55.238</td>
</tr>
<tr>
<td>Road Markings * Points</td>
<td>5.899</td>
<td>2.108</td>
<td>5.375</td>
</tr>
</tbody>
</table>

Table 3 Statistical analysis results for the curve Masseik

<table>
<thead>
<tr>
<th>Independent factor</th>
<th>F value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road Markings</td>
<td>10.568</td>
<td>.000</td>
</tr>
<tr>
<td>Points</td>
<td>6.636114.454</td>
<td>.000</td>
</tr>
<tr>
<td>Road Marking * Points</td>
<td>2.279</td>
<td>.000</td>
</tr>
</tbody>
</table>

Test of With-in Subject Effects (Greenhouse-Geisser)

<table>
<thead>
<tr>
<th>Independent Factor</th>
<th>Speed</th>
<th>Acceleration/Deceleration</th>
<th>Lateral Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road Markings</td>
<td>6.265</td>
<td>5.734</td>
<td>12.955</td>
</tr>
<tr>
<td>Points</td>
<td>119.579</td>
<td>78.543</td>
<td>45.798</td>
</tr>
<tr>
<td>Road Markings * Points</td>
<td>3.730</td>
<td>3.195</td>
<td>5.463</td>
</tr>
</tbody>
</table>
4.1 Mean Speed

Figure 2a and 3a show three speed profiles across the 11 points for all three conditions for the curves Hoogstraat and Masseik respectively. For Hoogstraat, independent variables points and the two-way interaction were found significant (Greenhouse-Geisser \( p\)-value\(<0.05\), Table 2) whereas road markings turned out to be insignificant for speed (Greenhouse-Geisser \( p\)-value\(=0.815\), Table 2). For Masseik, all three independent variables had significant effect on speed (Greenhouse-Geisser \( p\)-value\(<0.05\), Table 3). Drivers started to reduce their speed from the point ‘500MBC’ for all three conditions for both curves. The reason for this is that the curve was made visible approx. 500 meters upstream and a warning sign was placed 500m before the curve. Speed decrease before start of the curve was maximum for optical circles. This was expected as the objective of surface treatments is to reduce the speed of the driver before entering the curve.

4.2 Mean Acceleration

Table 2 and 3 show that all three independent variables were significant for acceleration at both curves (Greenhouse-Geisser \( p\)-value\(<0.05\)). Figure 2b and 3b show the plots of mean acceleration values for all three treatment conditions across the 11 points. In case of Hoogstraat, the acceleration in the optical circles case drops from ‘500MBC’ to the minimum value at ‘SOT’. The decrease in the acceleration was gradual, which also correspond to a second order (much smoother) change in speed for optical circles case compared herringbone and control cases. Increase in acceleration through the curve was largest for the herringbone treatment. This is because, for the herringbone treatment, drivers did not have to focus on correcting their lateral position and only required to focus on speed and acceleration. From this, we can assume that optical circles had a positive influence on speed and acceleration as they were able to reduce speed gradually. Speed did reduce for control and herringbone pattern but acceleration values suggest that speed before entering the curve was not decreased gradually rather abruptly.
Investigation of Herringbone pattern and Optical Circles for Safe Driving Behaviour at Curves Using Driving Simulator

4.3 Mean Lateral Position

For Hoogstraat, Table 2 shows that points and the two-way interaction between points and road marking are significant (Greenhouse-Geisser p-value < 0.05) for the lateral position of the drivers. For Masseik, Table 3 shows that all three independent variables were significant for the lateral position (Greenhouse-Geisser p-value < 0.05). Lateral position values for Masseik were found to be lower than they were in case of Hoogstraat due to narrower lane width. For both curves, drivers started to adjust their lateral position from the ‘SOT’ point, which is approx. 300 meters before the start of the curve, by shifting towards the right edge of the lane when herringbones were applied. For herringbone strip, drivers lateral position through the curve was approximately in the middle of the lane. However, for the control case, drivers were found to drive more towards the left edge of the lane through the curve. This might be considered unsafe as this can increase the risk of head-on collision with the traffic on the opposing lane.

5 DISCUSSION

In order to ensure safe driving through the curve, speed reduction should take place before drivers enter the curve as speed reduction in the curve can cause skidding of the vehicle which increases the probability of accident occurrence. Based on speed difference values between the points ‘EOT’ and ‘SOC’ for both curves, it can be inferred that both road marking treatments were able to reduce the speed of the drivers before entering the curve. Based on the concept of relative validity of the driving simulators, it can be assumed that the magnitude of the change in speed might be different in reality if same treatments are applied before and at the curve but would be in the same direction i.e. speed will decrease.

For optical circles, acceleration values decreased uniformly because the optical circles were applied 100m before the curve. Acceleration values for herringbone and control condition decreased sharply before the point ‘SOC’. This shows that optical circles were effective in safe reduction of speed before entering the curve. Though mean acceleration magnitude for all conditions was less than the recommended rate of -0.85m/s² (Lamm and Choueiri 1987), it can be assumed that variations can be expected in real life. For herringbone pattern, acceleration values started to increase after the point ‘SOC’. This is because drivers’ lateral position was controlled by the herringbone strips. As a result, drivers were comfortable to drive at higher speeds through the curve. For the lateral position, herringbones pattern’s influence was significant in the curve for both Hoogstraat and Masseik. The reason for this is the path drivers have to follow along the herringbone pattern is created in a way that the radius of the driver’s trajectory is increased. In case of optical circles and no treatment, drivers were found driving...
Investigation of Herringbone pattern and Optical Circles for Safe Driving Behaviour at Curves Using Driving Simulator

towards the inner edge of the lane which might increase the risk of head-on collision with the traffic in the adjacent lane hence, can be considered unsafe. Ariën et al. (2017) found that for speed reduction, transverse rumble strips were effective than the reverse herringbone pattern whereas for lateral position, the treatments did not had significant effect. In our study speed reduced for both treatments but acceleration values suggest that optical circles will cause a safe reduction in speed before the curve. Herringbone pattern was found to influence the lateral position of the drivers before and through the curve. This may be due to the appropriate design of the herringbone strips followed in our study.

6 CONCLUSION AND FUTURE RESEARCH

Results obtained for driving behaviour parameters show that both optical circles and herringbones had positive effects on driving behaviour. The optical circles caused safe speed reduction before entering the curve which makes them a more suitable option than herringbone pattern. However, for lateral position, herringbone pattern made drivers follow a safe path along the curve. This shows that herringbone pattern can significantly reduce the number of accidents on the curves where accidents occur mostly due to faulty lateral position of the drivers. Hence, it can be concluded that at curve sections where speed reduction is required, optical circles are better option whereas herringbone pattern is useful when lateral position needs to be controlled. Real world implementation of both treatments with before and after studies can allow policy makers to study the long term effects of both treatments. Moreover, comparison of different perceptual treatments at curves among themselves and with transverse rumble strips using a driving simulator may also be investigated in future.

REFERENCES

HMI DESIGN, HOW TO GET IT RIGHT?

Parallel session
Augmented Reality as an Advanced Driver-Assistance System: A Cognitive Approach

Lucas Morillo Méndez, CTAG, Spain, l.morillo.lm@gmail.com, Eva García Quinteiro, CTAG, Spain

ABSTRACT

AR is progressively being implemented in the automotive domain as an ADAS system. This increasingly popular technology has the potential to reduce the fatalities on the road which involve HF, however the cognitive components of AR are still being studied. This review provides a quick overview of the studies related with the cognitive mechanisms involved in AR while driving to date. Related research is varied, a taxonomy of the outcomes is provided. AR systems should follow certain criteria to avoid undesirable outcomes such as cognitive capture. Only information related with the main driving task should be shown to the driver in order to avoid occlusion of the real road by non-driving related tasks and high mental workload. However, information should not be shown at all times so it does not affect the driving skills of the users and they do not develop overreliance in the system, which may lead to risky behaviours. Some popular uses of AR in the car are navigation and as safety system (i.e. BSD or FCWS). AR cognitive outcomes should be studied in these particular contexts in the future. This article is intended as a mini-guide for manufacturers and designers in order to improve the quality and the efficiency of the systems that are currently being developed.

Keywords: Augmented Reality, Automotive, Cognitive, Review, Taxonomy.

1 INTRODUCTION

In the last years, Augmented Reality (AR) has been implemented progressively into the automotive industry in the form of Advanced Driver-Assistance Systems (ADAS). This technology is being employed in new vehicles as an aid for drivers in different circumstances such as navigation or hazard detection. It is expected it will help to reduce Human Factor (HF) related accidents on the road, but the relation between the use of AR during driving and its cognitive aspects is a relatively new topic that is still under development. This paper provides quick review of cognitive aspects of AR and a taxonomy of previous literature to date that makes special emphasis on the perceptual and cognitive aspects involved in AR while driving and its different uses. We consider that this review was necessary as a quick guide for automotive manufactures and designers: It is essential to explore the cognitive aspects involved in AR during driving in order to optimize the design of the technology, but it is even more important not to make things more difficult for drivers when developing a system. It could happen that the system offers little help to users instead of making things easier for them, adding extra risks in the form of cognitive and perceptual variables that were not taken into account during the design.

SCOPUS database was used as search motor in combination with manual web search. Used keywords were “Augmented reality” and “Automotive”. Over 173 results were displayed in the SCOPUS database after a first filter that excluded articles not related with HF. Over 50% of the results were related with the use of AR in the car manufacturing industry, but not directly with AR as an ADAS. These results were excluded, with the exception of three examples in the provided taxonomy (Kelly Villota et al., 2017; Langley, 2016; Marsh &
Merienne, 2015). In addition, technical results that involved design strategies without focusing on the cognitive aspects of AR were not used. Four additional articles were identified by examination of reference lists and other sources (Chung, Pagnini and Langer, 2016; Eriksson et al., in press; Lorenz, Kerschbaum and Schumann, 2014; Zimmermann et al., 2018). In the end, 26 publications were classified as useful for the purpose of this review.

2 COGNITIVE AND PERCEPTUAL IMPLICATIONS OF AR INSIDE THE CAR

There are many cognitive and perceptual aspects of AR that should be taken into account while driving. In this section some of them will be shown.

One of the potential drawbacks of AR inside the car is that it can lead to cognitive capture if the AR area is overloaded with information (Pauzie, 2015). This may cause the driver to respond to the offered information rather than to the driving task and a problem of overreliance, and thus, that users engage in risky driving. It is because of this reason that AR should keep a reasonable amount of essential information on display and this should not be secondary or not driving-related. A mix of important information and secondary information may cause confusion in the driver, who may not know where to look (Ng-Throw-Hing et al., 2013). In a study from 2016, Wang and Soffker suggested different Head Up Displays (HUDs) designs that displayed information related with driving efficiency. While driving efficiency information could be important for many people, it could be an example of secondary information that should not be provided in a HUD.

AR information does not have to and should not be present at all times, because this could also lead to overreliance on the system (Gabbard, Fitch and Kim, 2014). In a 2016 study related with the usefulness of AR while navigating, Chung et al. noticed that participants who had the possibility of doing punctual choices while navigating were more engaged with the task and showed less distractions. This is a complementary solution against overreliance: to offer drivers specific cues during specific manoeuvres, but never telling them exactly what to do, since that is their responsibility and could affect their driving skills in the mid-term (Ng-Throw-Hing et al., 2013). AR is a helping system, but it is not a substitute of real life signals. Even if AR allows us to keep looking to the road while looking at the information, it can receive more attention than the road, turning what is supposed to be a helping system into a distraction. Information should be presented just for certain manoeuvres and not constantly in order to increase situation awareness. Showing the relevant information for specific events makes the driver more attentive to what is happening. On the other hand, if the user had constant information, it could also occlude important visual information.

The concept of Useful Field of View (UFOV) is also important in this context. When talking about AR, UFOV refers to the visual area where information can be extracted with a simple glance without head movements (Ng-Throw-Hing et al., 2013). This UFOV is in accordance with the tendency of AR to be displayed via HUD inside the car, which do not require head movements to be seen. Using a HUD would be enough for obtaining benefits from AR and allows to reduce the potential cost of full windshield AR, so it makes sense that this is general tendency in the market. Having a full AR windshield could force drivers to move their head and thus, increase cognitive distance between two separated sources of information. Nevertheless, there have also been proposals of full windshield AR (see Charissis, 2014, Hosseini, Bacara and Lienkamp, 2014). More information and references about cognitive aspects of AR as an ADAS can be found in figure 1.
Figure 1 – Taxonomy of previous literature related with AR in the Automotive Industry

3 DIFFERENT USES OF AR INSIDE THE CAR

3.1 Safety

AR is a worthy system for providing safer driving. It can alert the driver and provide cues about the environment with enough time for them to react. Nowadays some cars include a Blind Spot Detection system (BSD), which informs the user about close cars or obstacles that they cannot see directly. Especially in these critical moments, it is crucial to remember that the way in which the information is presented to the driver can make the
Augmented Reality as an ADAS: A Cognitive Approach

difference. Cognitive implications of AR in the car specially apply here in order to optimize drivers’ cognitive resources and avoid fatalities. An example of safety system that have been studied in the AR domain is the Forward Collision Warning Systems (FCWS) (Yoon, Kim, Park, Park and Jung, 2014). In 2016, Phan, Thouvenin and Frémont designed an evaluated and tested a particular kind of FCWS, a Pedestrian Collision Warning System (PCWS). In their study, they found that conformal cues (adapted and in coherence to the real shape), in contrast with non-conformal cues (non-changing opaque shape), enhanced drivers awareness of the pedestrian.

3.2 Navigation

Another particular use of AR is navigation. Non-AR navigation process already requires a high cognitive workload for the user: visual attention and spatial orientation are cognitive functions that are highly involved in this context. The problem of navigation systems is that they also challenge the divided attention of the user. Divided attention is also known as multitasking, it is the ability of the person to perform two parallel activities at the same time. A straightforward way to reduce cognitive workload is offering the navigation information to the user directly in the HUD. In a study by Chung et al. (2016) it was suggested that AR helps the user to focus on while navigating. Several participants had to find a goal in a building with challenging routes. The authors showed that participants who used AR cues reached their final destination before the other group (no AR cues), supporting the use of AR for navigation. The use of AR for optimizing the navigation is also supported by the number of times the participants made mistakes (got lost), which was lower on the AR. It is important to notice that this study was done by pedestrians, and not drivers.

It has also been suggested that the AR navigational information should be provided only with 3D HUD because it allows to a better depth perception (Bark, Tran, Fujimura and Ng-Throw-Hing, 2014). In a different study (Pfannmüller, Krammer, Senner and Bengler, 2015), non-contact-analog (image not implemented over real environment) 2D screens were classified as more annoying and distractive than other 3D concepts. In that same study, the authors concluded that a pattern of small boomerang-like arrows was the ideal for navigation in contrast to a continuous arrow symbol. They argue that the second may occlude the real environment to a higher extent, and thus, lead to cognitive capture.

3.3 Take Over Request Assistant

Another promising use of AR is to assist drivers in a Take Over Request (TOR) event in the context of automated driving. Recent studies have shown that although AR does not necessarily decrease reaction times when regaining manual control, it does increase the quality of the TOR and assist drivers in decision making situations, like overtaking or braking when a car ahead is driving slower (Lorenz et al., 2014; Eriksson et al., in press). Combined, AR and autonomous driving systems could be a big step towards a much safer driving.

4 Conclusions

This paper is focused on the cognitive outcomes of interacting with AR while driving. Even though these systems are meant to help drivers, there is still work to do in order to fully understand how to make the best of them, and what is more, to not make these more dangerous than driving without them.

The information offered via AR should be essential and should make sense in the context of the real driving, this
Augmented Reality as an ADAS: A Cognitive Approach

means that it should be related with the main driving task. AR information should not be present at all times, avoiding occlusion of the road elements and over trust on the system. It is recommended that AR information is related with concrete maneuvers or particular moments.

Effectively used, AR may reduce accidents on which human factor is crucial. It is critical to keep researching the interaction between the driver and the AR device in order to design systems that take into account all the possible cognitive outcomes in the most efficient way.

REFERENCES


Bark, K., Tran, C., Fujimura, K., & Ng-Throw-Hing, V. (2014). Personal Navi: Benefits of an Augmented Reality Navigational Aid Using a See-THru 3D Volumetric HUD. In Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 1:1–1:8). New York, NY, USA: ACM.


Charissis, V. (2014). Enhancing human responses through augmented reality Head-Up Display in vehicular environment. In 2014 11th International Conference Expo on Emerging Technologies for a Smarter World (CEWIT) (pp. 1–6).


Langley, A., Lawson, G., Hermawati, S., D’Cruz, M., Apold, J., Arlt, F., & Mura, K. (2016). Establishing the
**Augmented Reality as an ADAS: A Cognitive Approach**

Usability of a Virtual Training System for Assembly Operations within the Automotive Industry. Human Factors and Ergonomics in Manufacturing & Service Industries, 26(6), (pp. 667–679).


Ipsilateral Versus Contralateral Tactile Alerts for Take-Over Requests in Highly-Automated Driving

Guy Cohen-Lazry, Ben-Gurion University of the Negev, Israel, guy.lazry@gmail.com, Nuphar Katzman, Ben-Gurion University of the Negev, Israel, nuphar@post.bgu.ac.il, Avinoam Borowsky, Ben-Gurion University of the Negev, Israel, Tal Oron-Gilad, Ben-Gurion University of the Negev, Israel

ABSTRACT

One of the most significant concerns regarding the use of Highly-Automated Driving (HAD) is drivers’ ability to regain control. Vibro-tactile alerts were already suggested as an effective modality for Take-Over Requests (TOR). However, it is not clear whether such alerts should be ipsilateral or contralateral relative to the location of hazards. Studies regarding tactile directionality in other domains, as well as in non-HAD vehicles have found mixed results.

In the current study, 15 participants drove a highly-automated vehicle in a desktop configuration driving simulator. Each participant experienced two TORs in which they were required to regain control and divert the vehicle away from an impending hazard, situated 4 seconds in front of them. The disengagement of the autonomous driver was signaled to drivers using a directional tactile alert. For half of the participants, the tactile alert was directed to the direction of the hazard (contralateral), for the other half, the alert was directed away from it (ipsilateral).

Results showed that drivers in both groups made the same amount of errors (initially steering the vehicle in the direction of the hazard before steering it away). However, when using ipsilateral alerts, drivers were faster to steer the vehicle away from the hazard. While this result is in contradiction to previous studies regarding the use of directional cues in driving, it is in-line with research regarding directional responses in other domains. We suggest an explanation for this discrepancy and discuss its implications.

Keywords: Highly-Automated Driving (HAD), Collision avoidance, Vibro-Tactile cues.

1 INTRODUCTION

Every day now, autonomous vehicles are being tested on streets in cities worldwide, and mass production of such vehicles is expected within a few years. However, these initial models would probably not allow Fully-Automated Driving (FAD), but rather Highly-Automated Driving (HAD) which requires occasional interventions of the human driver (e.g., Segal, 2017). This highlights the issue of transferring control between the automated and the human drivers, as has been pointed out by others before (e.g., Louw, Merat, & Jamson, 2015).

When using HAD, a vehicle may encounter situations where it reaches its design limits, or for some reason malfunctions. In such situations, the vehicle is expected to issue a Take-Over Request (TOR) to alert the human driver to take control. When reaching a system limit, it is expected that the vehicle would provide drivers with a sufficient amount of time to regain awareness before transferring the control (SAE International, 2014). However, system failures may cause the system controlling the vehicle to disconnect immediately (e.g., electrical failures in aviation incidents; for example AAIB, 2006). In such cases, the TOR should draw drivers’ attention to the road, and, preferably, provide them with immediate information regarding the driving situation.

Among the various possible modalities for alerting drivers about a TOR, vibro-tactile alerts provide significant
advantages, such as being gaze-free (Meng & Spence, 2015). Moreover, whether drivers are engaged in the driving task or a secondary task, it is more likely that they are using their auditory and visual senses, leaving the vibro-tactile channel clear for incoming alerts (Petermeijer, Winter, & Bengler, 2016). Previous research have found that vibro-tactile displays result in shorter brake times compared with visual or auditory displays (Scott & Gray, 2008). Nevertheless, vibro-tactile displays are not as effective in providing detailed information (Campbell, Richard, Brown, & McCallum, 2007) and their design must be carefully considered to achieve optimal results. Several studies have examined how adjusting display parameters and applying different designs affect drivers’ reaction times, performance and perceptions (e.g., Petermeijer, Cieler, & De Winter, 2017). Other studies focused on the location parameter, investigating the directional potential of vibro-tactile displays for drivers.

The location of vibro-tactile stimuli can be used to map lateral (Straughn, Gray, & Tan, 2009), longitudinal (Ho, Tan, & Spence, 2005) and even vertical information (Salzer, Oron-Gilad, Ronen, & Parmet, 2011). A tactile alert may be either contralateral (indicating the direction of a hazard) or ipsilateral (indicating the direction to steer to avoid the hazard). In non-HAD vehicles, the use of contralateral designs resulted in shorter reaction times (Wang, Pick, Proctor, & Ye, 2007), and was thus recommended as a guideline for the design of in-vehicle vibro-tactile alerts (Campbell et al., 2007). Surprisingly, this advantage of contralateral designs is contradictory to the principal of Stimuli-Response (SR) compatibility (Fitts & Seeger, 1953) according to which people react faster and more accurately when the stimuli and response are compatible (see Proctor & Vu, 2006). Based on the SR compatibility principle it would be expected that drivers react faster when an alert coming from a particular side would require them to steer towards that side, however, this is not the case. Müßeler et al. (2012) suggested that in a dangerous situation, people learn to avoid stimuli and thus to react in the opposite direction. Straughn, Gray and Tan (2009) make a similar claim, proposing that drivers “have learned to turn away from ‘naturally occurring’ warning signals such as a car horn or the sound of a collision” (p. 112). Another explanation is that drivers do not react solely based on the alert, but instead use it to assess the situation and react accordingly (Petermeijer et al., 2016) thus pointing them to the direction of the hazard reduces the time required to make the decision. Either way, both explanations are based on drivers’ awareness of their environment. However, researchers have already suggested that drivers of highly-automated vehicles quickly become disengaged from the driving task (e.g., Jamson, Merat, Carsten, & Lai, 2013). As a result, it is not clear whether they would react to a directional vibro-tactile alert in the same way drivers in manual vehicles do. As drivers become disengaged from the driving task and delegate the control and responsibility to the automated vehicle, they might revert to SR compatible reactions, rather than to driving-contextual reaction. The current study addresses this issue. Specifically, we tested whether the superiority of contralateral alerts in manual driving would apply to TOR alerts or whether, due to drivers’ disengagement from the driving task and the SR compatibility principle, ipsilateral alerts would result in shorter reaction times.

2 METHOD

2.1 Participants

Seven females and eight males, undergraduate students, aged 22-27 ($M = 23.9$, $SD = 1.57$), all having a valid driving license, took part in the research. Participants received bonus course credit for their participation and were free to withdraw from the study at any time.
2.2 Apparatus

Simulator. The research was conducted at the Driving simulator in the Human Performance Evaluation Lab (HPEL) at Ben-Gurion University of the Negev (see Figure 1a). The simulator consists of a 90 degrees display using three 24-inch computer-screens, running a simulation software to illustrate the driving environment (Realtime Technologies Inc. [RTI], Royal Oak, MI). Participants were seated about 1.1 m away from the screens.

Tactile interface. The tactile system consisted of a tactor controller (Eval2.0; Engineering Acoustics Inc. [EAI]) regulating six EAI-C2 tactors stitched to the car seat. The tactors (three on each side) were positioned along the exterior part of the seat, adjacent to participants’ thighs (see Figure 1b). The tactile alert was designed as three consecutive pulses (pulse duration: 250 milliseconds, pulse intervals: 450 milliseconds) This signal is perceived as urgent (Van Erp, Toet, & Janssen, 2015).

Secondary task. Whenever the automated vehicle had control, participants played a game of Simon to simulate the engagement in a secondary, non-driving-related task. For this game, an array of X colors appeared on the screen. The system presented a sequence of random colors for the participant to repeat (starting with two). Whenever participants repeated the sequence correctly, they were presented with a one-step longer series of colors (three, then four, etc.). If participants made an error, the sequence started again with two colors. The game is visually and cognitively demanding, thus it has higher potential in disengaging drivers from the driving task.

2.3 Driving Scenarios

Participants experienced two different scenarios. In each scenario, a malfunction in the automated vehicle was simulated and participants were required to regain manual control of the vehicle. At the time of the TOR, a materialized hazard (e.g., a vehicle blocking one of the lanes ahead) was situated 120 m ahead of the vehicle, blocking participants’ lane and other lanes either to the right of it or left of it. The participant was not able to see the hazard prior to the TOR. Participants had to steer the vehicle away from the hazard, as they did not have enough time to brake and stop the vehicle.

2.4 Experimental Design

Each participant experienced two TORs presented through a directional tactile alert. Participants were randomly assigned to one of two experimental groups. For half of the participants, the tactile alert was directed toward the hazard (contralateral group), whereas for the other half, the alert was directed away from it (ipsilateral group). Presentation time gap between the two succeeding signals was not fixed but did not drop below 3 minutes, to minimize the influence of the previously perceived tactile alert.
Ipsilateral Versus Contralateral Tactile Alerts for Take-Over Requests in Highly-Automated Driving

2.5 Procedure

Prior to driving the simulator, participants were introduced to the simulator and drove it both manually and in autonomous mode. Additionally, participants were familiarized with the vibro-tactile alert and its directionality as well as with the Simon game and its operation.

2.6 Dependent Variables

Two dependent variables were used. First, to examine whether using different directionality results in better decisions, we tested whether drivers initially steered towards or away from the hazard. The value for this binary variable was either Correct (i.e., first steer away from hazard) or Incorrect (i.e., first steer towards the hazard). Additionally, participants’ reaction time was measured as the time between the TOR and the first steering input. A steering input was defined as a 2-degree steering wheel change, as smaller values may be attributed to vehicle stabilization rather than to voluntary inputs (see Gold, Damböck, Lorenz, & Bengler, 2013).

3 RESULTS

3.1 Direction of first steering reaction

Our results show that the error rate did not differ between groups ($p = .509$; Fisher’s exact test), and was near chance level. Participants in the contralateral group made seven correct decisions out of 14, while participants in the ipsilateral group made nine correct decisions out of 16.

3.2 Response time to the first reaction

The small sample size in the current study did not allow the use of mixed ANOVA to test our hypotheses. Instead, groups were compared using independent-samples $t$-tests. Response time for participants in the ipsilateral group ($M = .82$ s, $SD = .22$) was significantly shorter than response time in the contralateral group ($M = 1.23$ s, $SD = .39$), $t(28) = 3.38, p < 0.01$, Cohen’s $d = 1.28$. Additionally, we examined the correct response time which was defined as the time between the initiation of the TOR and a 2-degree steering in the correct direction. Again, drivers in the ipsilateral group ($M = .99$ s, $SD = .35$) had a significantly shorter reaction time than drivers in the contralateral group ($M = 1.38$ s, $SD = .50$), $t(28) = 2.41, p < 0.05$, Cohen’s $d = .88$. An examination of response time differences between the scenarios revealed that drivers shortened their reaction times from the first to the second scenario (see Figure 2). Nevertheless, drivers in the ipsilateral group exhibited shorter reaction times in both scenarios.

![Figure 2. Response times for drivers in both groups, for the first and second scenarios](image)

* $p < .05$  ** $p < .001$
4 DISCUSSION AND CONCLUSION

The aim of the current study was to examine the use of directional vibro-tactile alerts for TORs in autonomous vehicles. Specifically, we examined whether ipsilateral or contralateral results in improved drivers’ performance. Our results show that in-line with the SR compatibility principle, when drivers are required to regain control of a vehicle after a TOR, ipsilateral alerts lead to faster reaction times. These results are in contradiction with previous driving-related studies (e.g., Wang et al., 2007). Nevertheless, similar results were found in other domains (e.g., aviation; Salzer et al., 2011) or when the context was neutral (Proctor & Vu, 2006). When this discrepancy was previously investigated in driving-related studies, most explanations regarded drivers’ context-awareness which led them to regard a vibro-tactile alert as a threat, and therefore to turn away from it (Straughn et al., 2009). These suggestions are based on the fact that drivers perceive the driving task as a threatening situation (Fuller, 1984). However, when being driven in an autonomous vehicle, and taking on a secondary task, drivers may become disengaged from the driving task and therefore also from its context and the sense of threat that relates to it. We suggest that by disengaging from the threat associated with driving a vehicle, when drivers are required to respond to a vibro-tactile alert they revert to the SR compatibility principle rather to the response more compatible with driving. This difference of contextual perception between drivers in autonomous and non-autonomous vehicles implies that when designing alerts and interfaces for autonomous vehicles, knowledge gained in non-autonomous vehicles must be used with care and adapted before it can be applied to autonomous vehicles. Our findings also reveal a significant improvement in reaction time between the first and second trials. In a future study, we intend to examine the effects of experience on drivers’ reaction times.

In conclusion, this experiment is a preliminary stage and provides a foundation for further examination in the field of directional vibro-tactile alerts in highly-automated driving. Our findings suggest that for TORs in highly-automated driving an ipsilateral alert may improve drivers’ reaction times. However, this claim should be examined after drivers’ gain system experience to further validate its significance for the design of TORs.

Acknowledgement Guy Cohen-Lazry is a doctoral student supported by a General Motors (GM Advanced technical Center) Inc. grant (GAC 2406) and Nuphar Katzman is a doctoral student supported by a grant from the US Army Research Laboratory through the GDLS subcontract, GDLS PO 40253724 (B.G. Negev Technologies and Applications Ltd) under Prime Contract no W911MF-10-2-0016 (Robotics Consortium), Robotics CTA 2015-2020, T2C1S3C, Michael Barnes, Technical Monitor. The views expressed in this work are those of the authors and do not necessarily reflect official policy.

REFERENCES


Ipsilateral Versus Contralateral Tactile Alerts for Take-Over Requests in Highly-Automated Driving

https://wwwnsjdrm01.storage.googleapis.com/MDgwNTgwMjgzNQ==01.pdf


Louw, T., Merat, N., & Jamson, H. (2015). Engaging with highly automated driving. To be or not to be in the loop. *In 8th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 190–196). Salt Lake City, Utah, USA.


**Lane Change Manoeuvres for Automated Motorway Driving Applications**

Christoph Klas, RWTH Aachen University, Germany, christoph.klas@fka.de, Nicole Eikelenberg, Ford Werke GmbH, Germany, Gudrun Voß, RWTH Aachen University, Germany, Lutz Eckstein, RWTH Aachen University, Germany, Maximilian Schwalm, RWTH Aachen University, Germany, Louis Tijerina, Ford Motor Company, USA, Peter Zegelaar, Ford Werke GmbH, Germany

**ABSTRACT**

To develop automated systems that can execute acceptable and comfortable lane changes on the motorway, it is useful to understand how lane changes are performed by human drivers, what factors influence the lane changes, and how they can be described parametrically. More specifically, it is important to understand the acceptable thresholds for parameters which are associated with the manoeuvre execution from a driver’s perspective. Furthermore, the cognitive process of the intention and decision for a lane change needs to be addressed to provide a well-accepted automated lane change decision and timing. This work aims at gaining first qualitative insights into these aspects. Two small-scale on-road studies were conducted to initially collect data on the process of human lane change behaviour in different driving scenarios. Objective data has been collected by equipping a measurement vehicle with a broad sensor set. Qualitative subjective data was gathered through different survey methods. In the first study, both the driver and a co-driver were included by means of questionnaires. In the second study, the ‘thinking aloud’ method was utilised to gain a deeper insight into the driver’s cognitive process. Based on the analysis of the objective data a parametric description of the lane change profile has been established. They can be used to generate automated lane change trajectories based on parameters from real data. For the evaluation of such automated lane change profiles, the implementation of driving studies in a high-dynamic driving simulator is useful and has been initially tested in an expert pilot study.

**Keywords:** Lane Change Manoeuvres, Automated Driving, Naturalistic Driving Studies, Driving Simulator.

**1 INTRODUCTION**

Automated motorway driving is a possible scenario for the introduction of automated driving features. Lane centring and adaptive cruise control systems combined already provide a basis for this use case. Adding automated lane changes would enable handling of most of the normal motorway driving situations. This paper describes some of the investigations aimed to identify automated lane change manoeuvre profiles for motorway driving that meet user requirements and expectations.

An important part of existing literature analyses lane changes is based on statistical investigations of naturalistic driving data (Fastenmeier, Hinderer, Lehning, & Gstalter, 2001; Hetrick, 1997). Many sources deal with mathematical modelling in order to recreate human lane change behaviour in traffic or driving simulations (e.g. Ehmanns, 2002; Kesting, Treiber, & Helbing, 2007). Others describe the use of these mathematical descriptions in automated driving functions (e.g. During & Pascheka, 2014). The data sources, level of detail and precision of the utilised measurement methods varied widely in these studies (Gipps, 1986; Hidas, 2002; Zheng, 2014).

For this work the lane change process has been structured into two phases (see Figure 1). The first is the decision
phase that describes the cognitive process which leads to the decision to perform or not perform a lane change including all necessary assessments of the traffic situation and rough planning of the possible actions. The second is the execution phase, in which the actual lane change with all preparation actions is performed.

2 ON-ROAD STUDIES

As literature and available data sets did not cover the linkage between comprehensive and detailed sensor data and subjective information, two on-road studies have been conducted to generate a data base of manually performed lane changes, including a qualitative subjective and objective data collection describing the manoeuvres.

2.1 First On-Road Study

Besides the generation of a first set of detailed objective data, the goal of the first study was to gain a first insight into the subjective decision process as well as comfort and acceptance rating of lane change executions. Also, possible differences in subjective ratings between the driver and the co-driver should be revealed.

2.1.1 Participants and Apparatus

In total, \( N = 10 \) subjects (\( n = 5 \) female) participated in the study. Half of them were drivers (mean age of 33 years; mean annual mileage \( \geq 15,000 \) km), the others were co-drivers (mean age of 24 years; mean annual mileage \( < 5,000 \) km) that did not drive during the whole study. Co-drivers were included to capture the perspective of passengers, because in automated vehicles their perspective on lane change appropriateness is relevant for customer acceptance. Each drive a driver and a co-driver were matched together resulting into five pairs. The study was performed in unsupported driving mode in a VW Passat CC, since natural driving behaviour needed to be captured, where the driver is responsible for accelerating, steering and braking. The vehicle was equipped with a sensor setup to monitor and log the vehicle environment as well as ego-vehicle variables. Two mid-/long-range RADAR sensors were used to log the positions of the surrounding traffic in front and behind the ego-vehicle. The bigger angular aperture of the sensor’s mid-range detection area guaranteed sufficient coverage of close vehicles in the neighbouring lanes. The vehicle’s motion state and global position was measured with a high-precision inertial measurement unit in combination with a differential positioning system. Furthermore, the access to the vehicle’s CAN bus of the vehicle allowed for acquisition of signals which described the driver vehicle interaction, e.g. steering angle, turn signal usage and accelerator and brake pedal values.

Questionnaires were used to evaluate the qualitative subjective input regarding the decision for/against and the execution of lane changes during the test drive. The drivers answered questions about (a) the arguments and intention for/against a lane change decision; (b) aspects to which they paid special attention; (c) the decisive
reason to (or not to) perform the lane change. Co-drivers were asked about their agreement with as well as a reasons for/against a lane change decision. Furthermore, they answered questions about the execution of an occurred lane change (e.g. overall acceptability, sportiness, cooperation, and safety).

2.1.2 Procedure

The route had an overall distance of 134 km (about 30 km on Dutch motorway, remaining part on the German motorway) with an estimated driving time of 1.5 hours. If existing, the speed limit varied between 80 and 130 km/h. To increase the number of routing-based lane changes, drivers were requested by the experimenter to take exits and entrances along the course. The test drives were carried out at a low traffic density outside rush hours.

After a short introduction, the driver (driver’s seat), co-driver (front passenger’s seat) and experimenter (back seat) entered the vehicle. As soon as the driver entered the motorway, the experimenter announced the first lane change situation to be evaluated and asked the driver to memorise the central elements that influenced the driver’s decision for/against a lane change decision (ego-vehicle position, environment etc.). In the meantime, the co-driver was asked to complete the questionnaire. Afterwards, the experimenter started to interview the driver. The next relevant situation was specified subsequently.

2.1.3 Subjective Results

Overall, \( N = 107 \) lane changes were subjectively rated and qualitatively evaluated. Regarding the decision process, Table 1 contains the main categories of reasons for a lane change decision as well as examples of the sub-categories mentioned by drivers and/or co-drivers. While each sub-category of the six main categories was named at least once by the drivers, co-drivers mainly focussed on reasons originating from speed regulation. That is, co-drivers did not identify environmental factors (e.g. cooperation, traffic rules, overview etc.) in their decision for a lane change. This result suggests that, in order to make an automated driving system comprehensible for a passenger, it may be advisable to inform passengers about relevant other environmental information they might otherwise miss.

<table>
<thead>
<tr>
<th>Main category</th>
<th># of sub-categories</th>
<th>Examples of sub-categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed regulation</td>
<td>6</td>
<td>Maintaining velocity; Relative speed to other vehicles</td>
</tr>
<tr>
<td>Cooperation/Interaction</td>
<td>6</td>
<td>Faster vehicles approaching from behind; Go with flow of traffic</td>
</tr>
<tr>
<td>Traffic rules</td>
<td>4</td>
<td>Speed limit; Obligation to drive on the right</td>
</tr>
<tr>
<td>Strategy</td>
<td>7</td>
<td>Choice of route; Choosing faster lane</td>
</tr>
<tr>
<td>Space</td>
<td>1</td>
<td>Free target lane/sufficient space</td>
</tr>
<tr>
<td>Safety/Timing</td>
<td>6</td>
<td>Avoid hectic/risky manoeuvres; Good overview</td>
</tr>
</tbody>
</table>

2.2 SECOND ON-ROAD STUDY

The focus of the second on-road study was to gain further insight into the relevant aspects in the decision process.
Furthermore, an estimation for the overall timings of the cognitive process should be established. As a first approximation, this was again performed on a qualitative data level.

2.2.1 Participants and Apparatus

In total, \( N = 7 \) subject drivers (\( n = 2 \) female) with a mean age of 33 years and a mean annual mileage of 10,000-15,000 km participated in the study. The same test vehicle with the identical sensor setup was used as in the first on-road study.

2.2.2 Procedure

A 120 km route on a German motorway was chosen to focus on higher relative speeds on German motorways. If existing, the speed limit varied between 100 and 130 km/h. Time slots for the test drives were shifted at 8 AM or 3 PM to get to get into denser traffic conditions. Again, random exits were chosen by the experimenter in order to add routing considerations.

In contrast to the first study, this study was conducted with only the driver and an experimenter on board. In order to gain insights into the driver’s cognitive process, the “thinking aloud” method was used. The drivers were encouraged to comment on all aspects which were relevant for their positive or negative lane change decision. They were asked to try to denote the point in time, where they feel, that a lane change will be necessary, to understand the time horizon of the lane change intention. The participants were asked to verbalise all their thoughts and considerations on their strategic and tactical actions with regard to their lane change or lane keeping behaviour using the categories of reasons for/against lane changes identified from the first study (see Erreur ! Source du renvoi introuvable.).

2.2.3 Subjective Results

In total \( N=506 \) situational events have been analysed in this second study. The subjective results were mainly qualitative insights into the cognitive process and influencing aspects from the traffic environment. One main result was that the drivers could not precisely specify the point in time, when the intention for a lane change arises in their mind. This supports the assumption that the motivation for or against a lane change is a continuous monitoring process with a comparative weighting of the subjective advantages of the different lanes. In addition to that, the drivers mostly mentioned more than one category of reasons for/against lane changes for their decision. However, the most frequent aspect for their intention and decision was the keeping of their desired velocity (Category “Speed Regulation/Overtaking”), especially for lane changes to the left. For lane changes to the right, the reasons in most cases referred to the sufficient space on the target lane combined with the obligation to drive on the right on German multi-lane roads.

The drivers’ verbalisation of their consideration underlined the complexity of this cognitive process, especially with regard to the number of different aspects which are monitored, interpreted, and partially predicted. Some new aspects for the formation of an intention could be revealed. E.g., it could be observed that the drivers preview distance can be very high (~300-500 m) in traffic scenarios with low complexity. Especially for trucks, which have a known (lower) velocity, the estimation of their influence is based on the vehicle type. Near ramps, the intention for lane changes to the right was prioritised lower, because there is an increased probability of
slower or decelerating traffic before exit ramps. The influence of routing considerations is subject to high interpersonal and intrapersonal variances (e.g. cautious drivers change lanes earlier than more dynamic, experienced drivers). In summary, the “thinking aloud” method allowed for a driver’s conscious motivation to be revealed. This would not have been possible with naturalistic driving data in general.

3 OBJECTIVE DATA ANALYSIS

The two on-road studies provided an extensive objective data base of measurement data for further analysis of the objective characteristics of human lane change behaviour. Objective data of 1073 lane changes and 1442 km motorway driving by 12 experienced drivers has been collected from both studies and has been pre-processed for further analysis. Below some of the analysis results and the derived model-based process for trajectory generation are presented.

3.1 Driving Data Analysis

Definitions of the general time course of a lane change were established to prepare the data analysis of the lane change execution process. The most important period is the time of the main movement into the target lane. Its start is defined as the last time, when the vehicle is aligned within the original lane before crossing the lane marking (i.e., when the ego-vehicle’s heading angle is (close to) zero). The end is set to the first alignment point in the target lane (i.e. when the ego-vehicle’s heading angle reaches zero again). This led to a well-defined lane change duration, which allows good comparability. It can be observed that 85% of the manoeuvres are between 5.5 and 8 s (based on the definition above). There was no dependency of this duration from the longitudinal driving speed of the vehicle.

As basis for the implementation of a technical representation of human lane change execution, an assumption of a two-part process was proposed. In the first portion of the steering action the driver’s main goal is to leave the original lane with a certain maximum lateral velocity. Therefore, a higher lateral acceleration and quicker change in the movement away from the original lane is adjusted. The goal of the second part, after crossing the lane marking, is to align the vehicle at a certain lateral displacement in the target lane with zero steering angle. As this set of boundary conditions needs to be met, drivers tend to adjust lower change rates and lateral acceleration, which also leads to a slightly higher duration of this portion.

3.2 Reproduction of Lane Change Profiles

The necessary input parameters for the first portion of the execution is the distance to the lane marking at the start of the lane change manoeuvre and the time at which this point shall be reached. The second portion can then be parameterised by the intended overall displacement. Based on these parameters and conditions, the necessary steering angle of the trajectory can be optimised by using a model-representation of the vehicle’s lateral dynamics. Within the optimisation, maximum values of the lateral acceleration and jerk can be considered to account for the driver’s comfort requirements. Comparisons of generated trajectories with measurement data reveal good matching, especially in terms of the course of the lateral velocity and the lateral acceleration.
4 GENERAL DISCUSSION

The presented work strived to identify user-centred automated lane change manoeuvre profiles based on objective and qualitative subjective data from non-automated on-road studies. The on-road studies helped to gain first insights into the structure and parameters of the lane change process. The studies also showed that the decision process for or against a lane change manoeuvre is manifold. Drivers seem to consider a lot of factors in their decision, while passengers do not always take all of them into account. However, these studies were based on small samples and were mainly analysed on a qualitative level. Therefore, further studies with larger and balanced samples are needed to receive robust results on a quantitative data level.

The current analysis of the objective results allows for a technical representation and reproduction of human lane change execution. These outputs can be used to create first synthetic lane change trajectories for the assessment of automated lane change manoeuvres based on parameters from real data. For a controlled and reproducible evaluation of such automated lane change trajectories, the implementation of high-dynamic driving simulator studies seems advisable. A pilot study in the high-dynamic driving simulator of the Institute for Automotive Engineering, RWTH Aachen University with $N = 5$ experts already revealed that the visualisation and the dynamics (unscaled simulation of actual dynamics) of the high-dynamic driving simulator are appropriate for future investigations of automated lane change profiles. The experts reported a realistic driving experience and that – based on their impression – drivers will probably prefer that the automated lane change is softer than a manual driven lane change. Further research will be conducted to verify and expand upon these preliminary findings.

REFERENCES


‘Trust me – I’m AutoCAB’: Using natural language interfaces to improve the trust and acceptance of level 4/5 autonomous vehicles

Vicki Antrobus*, Gary Burnett, David R Large
Human Factors Research Group, University of Nottingham, UK, *vicki.antrobus1@nottingham.ac.uk

ABSTRACT
A simulator study explored the use of a natural language interface (NLI) to improve the trust and acceptance of level 4/5 connected/autonomous vehicles. Twenty-three participants undertook three journeys, each lasting approximately 7 minutes. Journeys were framed as different ‘Taxi of the future’ scenarios, with the vehicle, serving as both the taxi and driver/assistant, transporting participants autonomously to a predefined destination. Participants were provided with one of two different interfaces (‘traditional visual’ and a NLI), designed to maximise passenger comfort, provide locally informed information and resolve journey-related problems encountered on route. The traditional interface communicated information visually using a touchscreen, whereas the NLI interacted with passengers using ‘natural’ spoken language. Interactions were limited to one type of interface for each journey, with experiences counterbalanced. For the purpose of the study, the NLI was created using a Wizard-of-Oz technique using an actor to mimic the natural language system. Given free-choice to select their preferred interface during the third drive, twenty participants chose the NLI. Subjective acceptance ratings revealed that participants considered the NLI as significantly more useful, also assigning significantly higher ratings of satisfaction and confidence compared to its traditional, visual counterpart. However, trust ratings were largely equivalent for both interfaces. Overall, results show that the NLI was preferred over traditional visual interfaces in an automated driving environment, supporting its use in future autonomous vehicles. Further work could explore how to adapt the linguistic interactions of natural language systems to also foster increased trust amongst users.

Keywords: Autonomous vehicles, Natural language assistant, Wizard-of-Oz, trust, acceptance.

1 INTRODUCTION
Trust has been described as an ‘individual’s willingness to depend on another party because of the characteristics of that party’ (Rosseau et al, 1998), and is considered to be a major factor in the acceptance of technology in the automotive domain (Lee and See, 2004) – ‘the degree to which an individual incorporates the system in his/her driving, or if the system is not available, intends to use it (Adell, 2009). The determinants of trust and acceptance in the context of humans’ interactions with future autonomous vehicles are thus likely to be complex and expected to derive from various factors, including the perceived usefulness and perceived ease of use (Davis et al., 1989).

Previous research has explored human-human interactions as a viable model for enabling future system communications, reasoning that as speech is the primary means of social interaction amongst humans, its use as a method of system interaction offers designers a unique opportunity to tap into human evolutionary instincts and foster intuitive interactions (Large et al, 2017). These evolutionary instincts are evident in some of our earliest behaviours. Born already primed to speech, we become able to distinguish speech like sounds from
our wider environment in early infancy. These skills are honed throughout our development, meaning as we grow we are increasingly able to extract socially relevant, salient and paralinguistic cues, even from limited exposure to speech, for example, making determinations about a speaker’s gender and personality from the pitch, cadence and rate of their speech, using these linguistic markers to guide our subsequent actions (Nass & Brave, 2005). These speech related behaviours endure even when we are consciously aware that the speech we are attending to has a non-human source. Designers have subsequently exploited these evolutionary behaviours, designing systems with varying ‘digital personalities’ which change their language and vocal characteristics to influence user trust, performance and learning (Nass & Brave, 2005).

Speech based interfaces have previously been examined as a combatant to fatigue (Large et al, 2018), a mediator of human emotion (Nass et al, 2005) and as a means to reduce the ‘gulf of evaluation’ between human and system (Eriksson & Stanton, 2017). These studies have shown that speech-based systems have the potential to improve driving performance and driver attitude when systems communicate using a voice consistent with the driver’s current emotion (energetic system voice for happy drivers and subdued voice for upset drivers), and indicate an emerging preference and increasing expectation for natural language interactions over other traditional interfaces in the automotive domain (Large et al., 2017). Additionally, Eriksson and Stanton (2017) suggest that the employment of natural language interfaces within the driving automation paradigm can enhance human-system communication by providing timely, intuitive feedback to drivers, ensuring they are aware of current system status, informing users why the system has chosen a particular path of action, and what the next projected action will be. The authors liken such system communications to a ‘chatty co-driver’, explaining that natural language can be used to relay information about vehicle sensor limitations and potential obstructions in the roadway ensuring that users are continually kept in the control loop should automation failures arise, thereby reducing the likelihood of incident. This study represents a first step in the exploration of natural language autonomous vehicle interfaces which specifically aim to foster trust and ultimately promote system acceptance.

2 METHOD

Adapting the driving simulator to mimic a level 4/5 automated vehicle, we utilised a Wizard-of-Oz approach to explore how interactions with a natural language interface (NLI) might improve the trust and acceptance in connected and autonomous vehicles, compared to a traditional, visually-based interface (as a baseline condition). Twenty-three participants (19 male, 4 female), with an average age of 28.7 years (SD = 8.69), completed three simulated ‘drives’, which were framed as a ‘Taxi of the future’. Each simulation took participants to a fictional destination (which was communicated to the participant prior to the journey), with the vehicle in autonomous mode and the participant seated in the drivers’ seat of the simulator. As such, participants were informed that they were not required to make any primary control inputs or engage with the driving environment during any stage of each drive. In their first and second drives participants were required to cooperate with system-initiated interactions from two different interfaces, each of these interfaces were framed as a ‘digital taxi assistant’, as follows:

1. “AutoCAB” - NLI which guided drivers though scenarios by providing participants with news updates, passenger comfort interactions (e.g. music selection), facilitating email and calendar organisation/
‘Trust me-I’m AutoCAB: Using natural language interfaces to improve the trust and acceptance of level 4/5 autonomous vehicles

scheduling, and finally offering assistance with emerging journey-related problems. Using a Wizard-of-Oz approach a male actor mimicked a natural language system via a video link within the driving simulator.

2. Traditional Interface - Used visual slides to offer matched assistance to passengers, ensuring that participants perceived both systems as having equivalent levels of functionality. Drivers interacted with the ‘system’ (with visual stimuli manually manipulated by the experimenter) using short verbal cues (e.g. ‘yes’, ‘no’, ‘more information’) akin to current voice activated systems within vehicles.

Participants were briefed of the capabilities of each interface and how best to interact with each system prior to commencing each simulation ‘drive’. In the third and final drive, participants were required to select which format of the digital assistant they would prefer to accompany them (AutoCAB or Traditional interface).

Participants completed trust in automation and system acceptance questionnaires after each simulated drive, which indicated their level of trust and acceptance of each assistance system. Ratings were elucidated by participants during post-trial interviews.

3 RESULTS

3.1 Acceptance data

After each simulator drive, participants were required to complete an acceptance questionnaire (Van der Laan, Heino & De Waard, 1997) which assessed system acceptance across two dimensions, perceived usefulness and satisfaction associated with system interactions.

![Figure 1 – Traditional interface (left) and Wizard-of-Oz setup (right) in which an actor mimicked an NLI interface](image)

![Figure 2 – Perceived usefulness and satisfying scores of participants after interactions with the ‘AutoCAB’ and Traditional interface (‘Interface’)](image)
A Wilcoxon signed ranks tests indicated that participants perceived the AutoCAB interface to be significantly more useful ($Z = -3.34, P<.001$) than the traditional interface. Similarly, participants also indicated that interactions with the AutoCAB interface were significantly more satisfying ($Z = -2.32, P<.020$) than the traditional interface.

### 3.2 Trust Data

Participants also completed a questionnaire which examined the level of trust between people and automated systems (Jian, Bisantz, & Drury, 2000). This assesses trust across a range of factors (Figure 3). A series of Wilcoxon signed ranks tests indicated both the AutoCAB and traditional interface were statistically equivalent across a number of aspects of trust. However, participants did report significantly higher levels of confidence in the AutoCAB (mean = 5.5) interface compared to the traditional interface (mean = 4.8) ($Z=2.43, p=0.015$).

![Participant Trust - AutoCAB/Traditional Interface](image)

**Figure 3 – Participant trust ratings for the AutoCAB NLI and Traditional Interface**

### 3.3 Choice drive

Overall results indicate that the natural language ‘AutoCAB’ interface was the preferred assistance system, with 20 out of the 23 participants selecting the NLI interface for their third, choice drive. Participants provided insight into the reasons for their interface preference in post-trial participants interviews. Those drivers who selected the AutoCAB stated that natural language communication was more intuitive, allowing them to communicate freely, without following defined system commands (P3). This free-flowing communication fostered feelings of trust in some participants (P11), meaning they felt more at ease when the car simulated a ‘system sensor failure’ as part of an experimental trust challenge. Additionally, interactions with the NLI system were perceived as being more ‘assisting’ (P13) and less onerous on the human passenger.

- **P11** – I prefer talking to people, I guess it comes down to feeling like you’re being heard. If anything happened like a breakdown I would want to be able to communicate that with system – If it was just the screen I would feel trapped in a box.
- **P13** – I felt that the first one (the NLI) was more trustable than the second. Maybe it was because I heard a human voice talking to me, I felt as if something was taking care of everything for me.

However, participant 12 did make the distinction that the social environment of the car could influence a person’s willingness to engage with an NLI system, offering that such systems such be adaptable, allowing the passenger to choose their level of interaction.
‘Trust me - I’m AutoCAB: Using natural language interfaces to improve the trust and acceptance of level 4/5 autonomous vehicles

P12 – I would prefer to interact via speech. Though, I guess it depends if you’re alone…. If you’re with someone, maybe not

Alternatively, those participants who selected to interact with the traditional interface in their third drive cited trust concerns (P2) and the potential sharing of information as reasons for their interface selection. However, this was not true for all participants selecting the traditional interface. Participant 6 suggested that their own social introversion meant that, any speech-based interactions were not preferred.

P2 – I don’t like voice based systems for trust reasons…. You don’t know what’s in the fine print (of systems) and what’s it’s recording…. It was disingenuous, it was obviously just using common points of conversation and using me as a data point.

P6 – I’m quite introverted. I don’t want to interact with people and machines – it’s not a trust thing.

4 DISCUSSION

The study compared the use of a NLI and a traditional visual interface to support drivers during automated driving, using their subjective ratings of trust and acceptance, and usage preference, as measures. Results indicated that overall, the natural language system was preferred to the traditional interface, inspiring higher ratings of usefulness, and confidence, and was found to be more satisfying; the NLI was also most commonly chosen during the ‘free-choice’ drive. Using natural language enables vehicle occupants to interact with the system without the need for training in specific verbal prompts or complex manual interactions, and reduces the visual demand associated with using the system, and this is likely to have influenced drivers’ preferences and ratings. Although participants were not actually driving, they would likely prefer to direct their visual attention towards secondary activities in a fully autonomous vehicle, and therefore minimising visual demand remains an important consideration. Additionally, the presence of a ‘human’ voice, and the conversational exchanges that ensued, is likely to have inspired perceptions of ‘humanness’ associated with the NLI. This ‘familiarity’ may have encouraged a more pleasant experience using the NLI, putting drivers at ease, and inspiring greater confidence that the system was ‘managing’ the journey effectively, as a fellow human might.

Nevertheless, some results did not fully support the aims of our study, i.e. there were no significant differences revealed in the level of subjective trust associated with the different interfaces. It is feasible that given the nature of the study design (i.e. with both experiences occurring in the simulator and in close succession), participants may have believed that both interfaces (NLI and traditional) were simply two manifestations of the same ‘system’, and as such, any differences were solely in the modality of presentation rather than in the underlying ‘intelligence’ of the system. Consequently, the information was considered to be equally ‘trustworthy’ in both situations. In contrast, the heightened confidence ratings associated with the NLI may be attributed to the paralinguistic cues that are naturally contained within speech and can modify the meaning and perception of vocal utterances. As such, participants may have drawn different interpretations of the information presented visually and orally, inspiring different ratings of confidence, even when the content of the utterances was the same. Additionally, heightened confidence ratings may be related to the participant’s ability to gain timely feedback on system status, and the ease with which they could interrogate the system. This is supported by Eriksson & Stanton (2017), who highlighted that clarity of system status through speech can enhance communication and individual perceptions of successful interactions.

Overall, the study provides some encouraging results that support the use of NLIs to increase the acceptance of future autonomous vehicles, although further work is required to understand the implications
for trust. In particular, future work should look at a wider range of use-cases (for example, employing ‘trust challenges’ to further test participants’ relationship with the system). In addition, the fidelity and validity of the ‘driverless’ experience could be improved, for example, by modifying the vehicle interior (removing primary controls etc. which would likely not exist in a fully autonomous vehicle, but were notably present during our study). In addition, further investigations could explore the linguistic properties and lexical content of utterances, with the aim of increasing trust using learned techniques from human-human interactions.

Acknowledgements: The research was conducted as part of the iMotors project, funded by Innovate UK (CAV 1) (ref: 102586).

REFERENCES


A “Driver-more” Approach to Vehicle Automation

Andrea Castellano, RE:Lab Srl, Italy, andreacastellano@re-lab.it, Serena Fruttaldo, RE:Lab Srl, Italy, Elisa Landini, RE:Lab Srl, Italy, Roberto Montanari, RE:Lab Srl, Italy, Andreas Luedtke, Offis, Germany

ABSTRACT

The scope of this paper is to describe an innovative interaction paradigm between the driver and highly automated vehicles, developed in AutoMate EU project. This new interaction modality is based on the cooperation, i.e. the mutual support in perception and in action between the driver and the car. The cooperation aims to exploit and make concrete the complementarity of the human and the automation.

The concept and the implementation have been tested by evaluating the system that more than others allows the driver to cooperate with the vehicle, i.e. the HMI. This tool has been selected since it enables the interaction, and indeed the cooperation, between the vehicle and the driver. The users were asked to interact with the HMI: a set of questionnaires was administered in order to test how this interaction paradigm is perceived and to measure the workload related to the interaction. The main impact of this approach is to increase the comfort and the acceptance of this disruptive technology, since the conceptual solution of mutual support can improve the relationship between these two agents, creating a third agent, the team, made of human-like technical solutions.

Keywords: HMI, Vehicle automation, Human-automation interaction, Negotiation-based communication, Cooperation.

1 INTRODUCTION

In the next few years, vehicle automation will change radically the conception of the driving task and the relation between the driver and the vehicle. Different approaches have been used to reach the full automation (Jian et al., 2015), taking into account the human factors related to the driving task (Cunningham, 2015).

At the moment, the mainstream approach is to reduce the role of the driver through the automation, chasing the so-called driverless approach. This safety-oriented paradigm seems to put limited consideration into the role of the human, replacing his/her role with the automation. Moreover, the traditional approach at vehicle automation is based on task distribution. The cooperative approach follows the principle of flexibility in task distribution; in order to achieve this flexibility, based on the current demands of the situation, on the capabilities and on the state (attention, readiness) and the capabilities of the human driver, as well as on the automation, in the joint driver-vehicle a close coordination between driver and automation is needed. The traditional, technological, approach is expected to transform advanced driver assistance systems gradually into fully autonomous cars.

But until the driver is in the loop (i.e. not fully autonomous), we must develop an interaction strategy that leverages on both the strengths of the automation and the driver to overcome their limits and cope with the
A “Driver-more” Approach to Vehicle Automation

complexity of the real world.

2 BACKGROUND

Two agents can be described as in a cooperative situation if they meet two conditions: (i) each strives towards goals and can interfere with the others on goals, resources, procedures, etc. Interference can take several forms, for example precondition, etc. If there is no interference, coordination is prebuilt and is not questioned during task execution; thus, the agents’ activities are independent; (ii) each tries to manage interference to facilitate the individual activities and/or the common task when it exists ((Hoc, 2000). Therefore, the cooperation happens when humans and machines execute tasks together, use resources together, interact with each other in a dynamic environment and make autonomous decisions. So, the cooperation is not only a matter of task distribution, since it concerns also a perceptual and decisional level. This concept, based on the Hoc model of cooperation shown in Errore. L’origine riferimento non è stata trovata. (Hoc, 2009) has been exploited and extended in the project’s framework, and it will be described in the next chapter.

The success of more complex and more automated vehicles in the future will depend on how well they interact, communicate and cooperate with humans both inside and outside the vehicle. The quality of cooperation and communication will strongly determine the driver’s trust in the automated systems. And in order to leverage the introduction of highly automated vehicles to the market and to fully exploit the automation potential to improve traffic safety and efficiency, these systems need to be trusted by the driver appropriately.

In terms of communication between humans and highly automated vehicles different solutions have been developed in the last years. For example, a novel and relevant solution implies multimodal communication. Multimodality in vehicle’s HMI can be considered from two different points of view, namely, multimodal input and multimodal output. A multimodal input describes the driver-to-vehicle information flux; whereas, a
multimodal output describes vehicle-to-driver one. Multimodal elements help to reduce the complexity of car’s cockpit. Nonetheless, each single interaction modality has advantages and drawbacks. These solutions that combine audio, visual interaction and sophisticated images, have been implemented in the TeamMate HMI. The paper aims at evaluating the concept of human-automation interaction based on cooperation. In particular, the technical enablers developed in this project have the scope to suggest an innovative interaction paradigm to overcome the concept of “driverless” approach to suggest a “driver-more” approach, i.e. to increase the role of the driver in the driving task instead of reducing it. In this way, the human and the automation will be seen as members of the team, and the cooperation is the means to compensate the reciprocal limits. It concerns mainly a different and unconventional use of technological supports to exploit the complementarity between human and technological agents.

This new approach implies the need to consider Human Factors implications related to the interaction with highly automated systems. For this reason, one of the scopes of this study is to understand if the greater involvement required to the driver compared to traditional systems is likely to increase the workload and affect his/her feelings, e.g. increasing frustration.

3 PROJECT CONCEPT

The top-level objective of AutoMate is to develop, evaluate and demonstrate the “TeamMate Car” concept as a major enabler of highly automated vehicles. This concept consists of considering the driver and the automation as members of one team that understand and support each other in pursuing cooperatively the goal of driving safely, efficiently and comfortably from A to B. In order to describe how the cooperation is actually implemented, it is important to briefly explain why the cooperation is needed, and how the human and the automation can support each other to create a safe, efficient and comfortable driving experience. Both the human and the automation have limits that can negatively affect the safety as well as the efficiency, the comfort, the trust and the acceptance of the autonomous driving. For the human, the limits are often related to his/her driving performance: they are likely to affect the safety, and cause accidents. For the automation, the limits, mostly at perception and decision level, may affect the efficiency and the comfort of the trip, and then, in turn, the acceptance of the automation. AutoMate goes beyond the simplistic concept of trusted automation by proposing a much more sophisticated approach to create trusted human-automation teams based on a new concept of cooperation. The AutoMate approach is based on the mutual complementarity between the driver and the automation: this support is achieved through the cooperation, between the team members.

While the Automation to Human Cooperation (A2H) is used to complement the human limits, the Human to Automation Cooperation (H2A) is implemented to allow the driver to support the automation to overcome its limits. The complementarity between the driver and the automation is the conceptual solution to compensate the reciprocal limitations, while the cooperation is how the complementarity is implemented. Figure 3 shows how both the A2H and the H2A cooperation can be implemented in perception (state A and B) and in action (state C and D).
To implement this disruptive concept, a shift of paradigm in the role of the driver is required. The technology should not decreasing the human element, yet it must support the driver to have a new role: he/she must be involved with a different role in the driving strategy. What is this new role? While increasing the levels of automation, the role of the driver shifts from driving tasks to decision-making tasks to complement the limits of the automation. In this scenario, the HMI is the key enabler to implement the innovative cooperative concept: a new concept of interaction has been designed to allow the automation to involve the driver in the driving strategy to deal with the complexity of the real world. The archetypal interaction paradigm, that is warning-based, is still used to suggest critical situations, but when the situation is not safety-critical the interaction is negotiation-based.

So, the HMI developed in Automate is innovative in terms of concept, since it introduces a new interaction paradigm, but is also disruptive in terms of implementation. The success of this strategies relies less on extraordinary intelligence and more on sophisticated negotiation of changing context and subsequent behaviour. For this reason, we used a 3D video to deeply explain the cause of the request of support and the expected behaviour. Both in terms of concept and implementation the most disruptive part is the H2A in perception because it allows to reduce the number of requests of takeover (a well-known safety critical condition in the driver-automation interaction).

4 TEST METHODOLOGY AND RESULTS

A test to evaluate the HMI was performed on the instrument cluster, since it can be considered as the main tool in which the cooperation happens. The study was preliminary and a small sample of users were involved. The number of participants selected for the test was 9. The gender of the subjects was balanced to avoid possible biases: 5 males and 4 females were recruited. The average age of participants was 29,44 years (Range=39-23; σ=4,55).
A “Driver-more” Approach to Vehicle Automation

Research questions for the specific H2A support (in perception compared to the support in action, i.e. the request of takeover from the vehicle to the driver) were:

I. Does the driver understand that the automation is asking for a support only in perception?

II. Does the driver understand that he/she has only to use his/her perception (in place of the sensors of the automation) and provide a feedback, instead of taking back the control of the vehicle?

III. Is the support in perception able to reduce the cognitive workload and the frustration related to the disengagement?

The participants were asked to interact with the HMI in a static experiment; they were invited to examine the HMI’s behaviour in different use case’s scenarios, and then to answer some questions to measure the level of comprehension of the message. The message was display in a digital instrument cluster and reinforced with audio, i.e. vocal communication in natural language from the vehicle to the driver, to adapt the communication to the complexity of the scenario. In order to measure the workload, the NASA-TLX questionnaire was administered after each scenario. This tool is used to measure different levels of workload, including mental, physical and temporal demand (Hart et al., 1986). It is a self-reported measure, widely considered as a solid technique to quantify the workload. It has been used in different domains, including aviation (Zheng et al., 2011), and automotive (Pauzie and Manzano, 2007), including autonomous vehicle research (Schipani, 2003).

The research questions (I) and (II) were stated as requirements to validate the HMI: results show that 96,3% of answers about the type of support needed were correct (success criteria: > 90%); (I) moreover, the message was correctly understood by the users, i.e. we’re able to understand that only a perceptual help was request by the vehicle (results: 96,3% - success criteria: > 90%).

The results of the NASA TLX questionnaire show that the support in perception is considered as less demanding then the support in action, confirming the hypothesis and giving strength to the approach established in the concept. In fact, the overall workload perceived by the users was lower for the support in perception than the support in action ($\Delta = 0,73$). In particular, the support in perception proved to be effective in improving the perceived performance ($\Delta = 1,05$), reduce the effort ($\Delta = 0,88$), and reduce the frustration ($\Delta = 1,25$).

Although the HMI for H2A support in perception (negotiation-based) is more complex than the warning-based HMI (to adapt the amount of

![Figure 4: Average workload in perception and in action](image)
information to the complexity of the situation), the users perceive less effort when the cause of the need (i.e. the limit) of cooperation is explained.

5 CONCLUSIONS AND EXPECTED IMPACT

This paper describes the concept of cooperation developed in AutoMate project, how it has been implemented through the HMI and tested to assess its consistence. In particular, the concept of cooperation, as an implementation of the complementarity between the human and the automation, has proved to be a promising solution to increase both safety and acceptability issues arising from the increasing vehicle automation. The results of the experiment suggest that the most innovative direction of support (i.e. H2A, when the automation requests a support to the driver) is well understood and accepted by the users. Moreover, the H2A support in perception has been measured to be less demanding then the support in action (the transition of control). Although this study is preliminary and has several limitations due to the small sample size and the lack of statistical analysis, this pointing can be considered a promising outcome. In fact, although the HMI for H2A support in perception (negotiation-based) is more complex than the A2H warning-based HMI (i.e. the archetypal paradigm used in automotive HMI industry and research), the users have proven to be able to understand it and to perceive correctly the reduced requested effort compared to the H2A support in action (i.e. the request of takeover). The findings emerged during this study are likely to be used as a basis to design effective Human Machine Interfaces for highly automated vehicles. This concept and its implementation can be considered an innovative solution that is likely to be implemented in a short-term period, before the adoption of fully automated vehicles. AutoMate project follows an iterative path; the next steps foresee to test this concept in more complex scenarios with a larger sample of users, in driving simulators and real cars.

So, why “driver-more”? Because to increase the involvement of the driver it means to foresee a near future of cooperation that does not weaken the driver, rather it empowers him/her, reinforcing the role of the human.

As stated by Ju, “future adopted automation systems will not wrestle control with the human: the team will negotiate activities, communicate and reconcile disparate perceptions of the environment and anticipate actions with explicit and implicit interactions” (Ju, 2015).

REFERENCES

A “Driver-more” Approach to Vehicle Automation


Koo, J. (2014), Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance, International Journal for Interactive Design and Manufacturing (IJIDeM) 9(4)


Schipani, S. P., An evaluation of operator workload during partially-autonomous vehicle operations, DTIC document


ACKNOWLEDGMENT

This research has been funded by the European Commission (H2020-MG-2014-2015) in the interest of the project AutoMate (http://www.automate-project.eu/) – Grant Agreement 690705.