What if it Suddenly Fails? Behavioral Aspects of Advanced Driver Assistant Systems on the Example of Local Danger Alerts

Angela Mahr[1] and **Yujia Cao**[2] and **Mariët Theune**[2] and **Veronika Dimitrova-Krause**[1] and **Tim Schwartz**[3] and **Christian Müller**[1]

Abstract. Many researchers argue, in assessing the benefits of Advanced Driver Assistance Systems (ADAS) it has to be taken into account that any gains in terms of security may be again reduced by the fact they affect the drivers' behavior. In this paper, we present results of a driving simulation study in which we compare driving performance as well as driver stress in three conditions: 1) when no assistance is available (NOASSIST) ; 2) when the system is working as it should (WORK); 3) when the system suddenly fails (FAIL). Results show that the driving performance is severely affected by system failures. The drivers' ability to effectively react to suddenly appearing obstacles FAIL is significantly lower than in NOASSIST. At the same time, the stress level is significantly higher in FAIL compared to WORK.

1 Introduction

The benefits of Advanced Driver Assistance Systems (ADAS) are potentially very large because they may considerably contribute to decreasing human suffering, economical cost and pollution [2]. However, as many researchers argue, any gains in terms of security may be again reduced by the fact they affect the drivers' behaviour [2, 6]. In this study, we investigate such behavioural changes. Particularly, we compare driving performance and stress 1) without assistance, 2) with help of the system, 3) and in a situation where it suddenly fails. According to [6], the following behavioral changes might occur: decrease or shift of attention, risk adaptation, and over-reliance. Over-estimating the functionality or reliability of the system might increase the effects of the former two: The sheer trust in the system, even if it is not fully justified, might make us drive riskier or become less attentive. Thus, over reliance is an important factor in the overall equation that results in negative effects on safety. The study presented here aims at investigating the effects of unanticipated system failures due to over reliance on driving performance and stress. Our basic assumptions are supported by signal detection theory [7]. It has been demonstrated, that an increase in false alarms decreases the operator's *compliance* resulting in longer response to or even disregarding of alerts [5]. An increasing miss rate, on the other hand, leads to a reduction of *reliance* and to closer examination of raw data in order to better avoid missing anything. Conversely, if during a longer period of time only a marginal percentage of misses is occurring, the driver might excessively trust the warning system and be less conscientious when checking the raw data or even rely completely on the system.

[1] German Research Center for Artificial Intelligence
[2] Human Media Interaction, Universtity of Twente
[3] Cluster of Excellence M2CI, University of the Saarland

2 Experiment

With over-reliance as the major factor to be investigated, we conducted a user study in which obstacle alerts were presented in various modes. While differences in presentation mode were reported in [3], we focus here on three conditions: 1. NOASSIST, which was always processed first; 2. WORK, which took the bulk of the experiment duration and therefore we expected, that the drivers would rely on the warnings; 3. FAIL, which was passed through only twice at the very end of the experiment. 32 subjects were paid to participate (16 men and 16 women between 20 years and 62 years, mean = 32.6, standard deviation = 10.8). The driving scene was a one-way highway with two lanes and no extra traffic projected onto the windshield. Visual warning messages were presented on a 10.6-inch head-down display. Auditory signals were delivered through a PC speaker located in the center of the vehicle. A warning message described an obstacle in terms of its type (what), location (where) and distance (how far). Drivers were required to change to the offside lane if the obstacle was on the nearside lane and brake if the obstacle was on the offside lane or on the right roadside. In a within-subject design, multiple experimental conditions with altered presentation mode were used in order to establish a sufficient level of reliance. In 16 % of the cases, false alarms (no misses) were interspersed in order to make the system behavior more realistic. At the very end of the experiment two additional obstacles were presented without warning (misses). These two obstacles became visible to the driver at a distance of 70m, which is identical to the setting in the NOASSIST condition. Measurements were derived according to ISO1998 usability model. Three error types regarding driving safety were distinguished: incorrect reaction, such as lane change instead of braking, late reaction, and no reaction. *Errors* measured the total amount of these three types of behaviors in each condition. The reaction time for NOASSIST and FAIL was defined as the time interval between the moment when an obstacle appeared and the moment when an action was performed. A brake action was identified when the speed changed from 120km/h to 60km/h (speed changed abruptly once a subject hit the brake pedal). A lane change action was identified when the lateral displacement of the car reached 10 % of the maximum lateral displacement during the course of a lane change. A low pitch error sound was delivered in case of a late or missed reaction. Additionally, we measured the stress using a high resolution skin temperature sensor. Skin temperature is an effective indicator for objectively evaluating stress, because it is controlled by sympathetic nerve activity which reflects the course of information processing in the brain [8]. Sudden drops in the curve were calculated and attributed to particu-

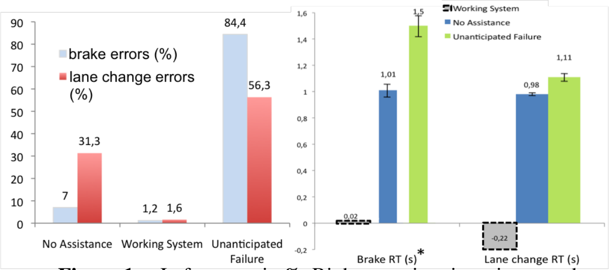

Figure 1. Left: errors in %. Right: reaction times in seconds.

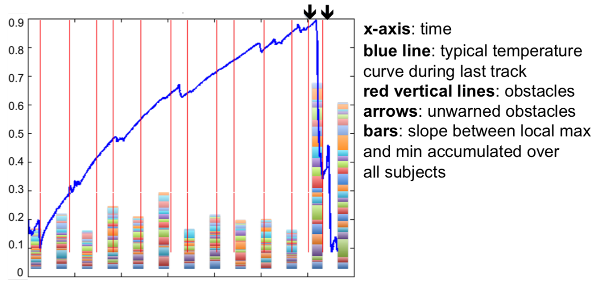

Figure 2. Results regarding stress on the basis of body temperature.

lar (announced respectively unannounced) obstacles. The functional value used for statistical analysis is $abs(delta(norm(temp)))$.

3 Results

Figure 1 (left) shows the percentage of not circumnavigated obstacles. On average, 19.1 % of the maneuvers failed in NOASSIST, while only 1.4 % failed in WORK. Sudden systems failures caused a drastic increase of unsuccessful reactions to 70.3 %. In NOASSIST, there was an error rate of 7,0 % for brake 31,3 % for lane change reactions, which indicates that the latter task was harder. In WORK, the percentage of errors is reduced to a marginal level for both type of responses. Regarding the two FAIL situations, it is important to keep in mind that subjects always encountered the braking situation first. The error rate of 84,4 % for braking decremented by almost 30 % absolute to 56,3 % for lane change (second unwarned obstacle) – after the initial surprise drivers apparently became aware of the possible system failure. Taking into account that lane change reactions were more difficult, the actual difference between first and second reaction can be assumed to be even greater than 30 %. In Figure 1 (right), reaction times are contrasted. Here, WORK has a clear advantage since reactions could already begin before the obstacles became visible (negative values). In NOASSIST, reaction times of all 32 subjects with respect to all 16 encountered obstacles could be averaged. FAIL on the contrary, is inherently a low-number-trial situation. For the first obstacle without warning it is crucial to notice, that 71 % of the subjects did not react at all. So the reaction time for this condition is derived from only those subjects who did (N = 9). Remarkably, there still can be found a significant difference in the brake reaction times for these 9 subjects, $t(8) = 5.8$, ($p < .001$), which is clearly indicating a delayed reaction. When approaching the second unwarned obstacle all subjects without exception tried to change the lane (N = 32). Nevertheless the reaction times of NOASSIST and FAIL differ significantly, $t(31) = 4.3$, ($p < .001$).

Figure 2 shows a typical (normalized) temperature curve during the final track of the experiment. The positive trend can be attributed to relaxation effects as well as a warming climate inside the car during a single track. The vertical red lines correspond to the time points when obstacles were presented. Sudden drops in the curve indicate increase of stress. The points marked with arrows represent the time points when the two unwarned obstacles were presented. Stress level drastically increased with sudden system failures. In the repeated case, the drop was not as severe, which confirms the findings based on driving performance. The bars represent the slope between local maxima and minima over all subjects. WORK was compared with FAIL in a repeated measures ANOVA, which confirmed a significant main effect of condition ($F(2,23) = 32.1$, $p < .001$). Helmert contrasts clearly revealed that skin temperature dropped significantly more when no warnings were present ($F(1,24) = 66.4$, $p < .001$).

4 Conclusions

The reliability of an automation system can be considered in terms of false alarm rate (FA) and miss rate (MISS). By varying the threshold settings for the decision criterion, designers are often able to trade one against the other [9]. Previous research has stated that FA could be more harmful to the user's performance than MISS [1, 4, 10]. A high FA might lead to annoyance, distrust and even ignoring of the valid system outputs. On the other hand, with a certain level of MISS, the automation can still be beneficial, especially in a demanding multi-tasking situation. If users are aware of the possible but relatively rare MISS, they would stay vigilant of the raw data (e.g. the road in our case). Thus, the misses might still be successfully detected. It was suggested that reliability levels of 70 % to 75 % represent an optimum threshold of imperfect reliability assistance [10]. Consequently, a 75 % reliable system would be the best to be used on the road. Our findings veer toward this argumentation as it is shown that very rare misses potentially lead to severe performance decline accompanied with a rising stress level. The findings reported here give rise questions on the roll-out strategy for fully autonomous cars, either into large-scale field test or into practice. The technology is susceptible for behavioral impacts such as attentional decrease/shift and transition problems in combination with over-reliance. According to our results, the behavioral effects have to be taken into account. Engineers should make sure that the drivers are always aware of the fact that the system may fail.

Acknowledgements. This work was funded by the Dutch Ministry of Economic Affairs (BSIK03024) and the German Ministry of Education and Research (01IW08004).

REFERENCES

[1] J. Bliss. An investigation of alarm related accidents and incidents in aviation. *Intern. Journal of Aviation Psychology*, 13:249–268, 2003.

[2] K. A. Brookhuis, D. de Waard, and W. H. Janssen. Behavioural impacts of advanced driver assistance systems – an overview. *European J. on Transportation and Infrastructure Research*, 1(3):245–253, 2001.

[3] Y. Cao, A. Mahr, S. Castronovo, M. Theune, C. Stahl, and C. Müller. Local danger warnings for drivers: The effect of modality and level of assistance on driver reaction. In *Proc. IUI 2010*, Hong Kong, 2010.

[4] S. R. Dixon, C. D. Wickens, and J. S. McCarley. How do automation false alarms and misses affect operator compliance and reliance aerospace systems: Cognitive factors in aviation. In *Proc. Human Factors and Ergonomics Society 50th Annual Meeting*, pages 25–29, 2006.

[5] S. R. Dixon, C. D. Wickens, and J. S. McCarley. On the independence of compliance and reliance: Are automation false alarms worse than misses? *Human Factors*, 49(4):564–572, 2007.

[6] M. Gründl. *Fehler und Fehlverhalten als Ursache von Verkehrsunfällen*. PhD thesis, University of Regensburg, 2005.

[7] J. Meyer. Effects of warning validity and proximity on responses to warnings. *Human Factors*, 42:563–572, 2001.

[8] C. K. Or and V. G. Duffy. Development of a facial skin temperature-based methodology for non-intrusive mental workload measurement. *Occupational Ergonomics*, 7:83–94, 2007.

[9] R. Parasuraman, P. A. Hancock, and O. Olofinboba. Alarm effectiveness in driver-centred collision-warning systems. *Ergonomics*, 40(3):390–399, 1997.

[10] C. Wickens and S. Dixon. The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science*, 8(3):201–212, 2007.