A SCHEMA OF POSSIBLE NEGATIVE EFFECTS OF ADVANCED DRIVER ASSISTANT SYSTEMS

Angela Mahr, Christian Müller
German Research Center for Artificial Intelligence (DFKI)
Saarbruecken, Germany
firstname.lastname@dfki.de

Summary: The purpose of Advanced Driver Assistance Systems (ADAS) is to enhance traffic safety and efficiency. ADAS can be considered as a (still incomplete) collection of systems and subsystems towards a fully automated highway system, such as autonomous cars. However, as many researchers argue, in assessing the benefits of 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 introduce a schema of possible negative effects of advanced driver assistant systems according to which consequences of a system failure largely depend on the magnitude of over-reliance. Based on that schema, we itemize hypotheses on possible behavioral effects of a specific ADAS type, namely local danger alerts.

INTRODUCTION

The purpose of Advanced Driver Assistance Systems (ADAS) is to enhance traffic safety and efficiency. Their benefits are potentially very large because they may considerably contribute to decreasing human suffering, economical cost and pollution (Brookhuis et al., 2001). Consequently, economic principles as well as an unacceptable number of accidents have given rise to an increasingly fast development of electronic driving aids. ADAS can be considered as a (still incomplete) collection of systems and subsystems towards a fully automated highway system, such as autonomous cars. ADAS may operate in advisory, semi automatic or automatic mode, all of which may have different impact on driving and consequently also on traffic safety.

However, as many researchers argue, in assessing the benefits of 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 (Gründl, 2005). In this paper, we review the literature on behavioral aspects of ADAS, adopt a set of criteria introduced by Pfafferoth & Huguenin (1991), and introduce a schema of negative effects. We furthermore introduce local danger alerts (LDA) as a particular type of ADAS and apply the set of criteria to it. Finally, differences between the experimental situations for local danger alert assessment and real life are discussed.

PROPOSED SCHEMA

First of all, providing information (e.g. warning messages) potentially leads to a situation where the driver's attention is diverted from traffic. This issue has to be taken into account when designing in-car systems. In the context of local danger alerts, for example, (Cao et al., 2010) compare various modalities (vision, speech) in order to find out the optimal way to deliver the warning message. However, we do not consider this effect a behavioral aspect as it represents an immediate reaction to the system rather than a side effect of taking over (parts of) the driving task. The latter might result in a reduction of the drivers' level of attention (Brookhuis et al., 2001) or make her/him engage in some other unrelated task causing driver distraction – from selecting music to reading a novel in extreme cases (Gründl, 2005). Let us call this behavioral effect "attention decrease/shift" (AD/S) as this term is addressing not only behavioral effects like "driver distraction" as an attention shift, but also a decrease of the overall driver's attention level. As a result, the driver might too late become aware of a sudden hazard or might be not ready for an adequate reaction. If we look at a sudden change between easy and difficult driving conditions (which is not necessarily a hazard), we can speak of a "transition problem" (TP). Here, the driver is not able to perform the sudden shift between cognitive underload (easy driving conditions with help of the system) to cognitive overload (difficult driving conditions eventually without help of the system). This effect also plays a role with fully autonomous cars since it is very likely that these systems will have to have a manual override mode in practice.

Another negative effect of ADAS might be drivers taking risks, which they would not take without the system. The likelihood and magnitude of such "risk adaptations" (RA) are not random or arbitrary, but rather depend on certain conditions. Pfafferoth & Huguenin (1991) introduce five general criteria that influence the occurrence of RA in the context of security measures. According to the authors, the likelihood of RA increases

1. the more the driver interacts with the system,
2. the more immediate the feedback is,
3. the more the system widens the driver's scope of action,
4. the more it increases subjective safety,
5. and the more it superimposes risky driving tendencies.

Hence, in order to adapt to the new condition, the driver has to be aware of the system or at least needs to know that it exists. Assistance systems that the driver is not aware of until a (near) crash are unlikely to cause RA effects. Consequently, RA becomes more likely if the driver is able to gain concrete experience with the causes and effects of the system by interacting frequently and immediately with it. Obviously, autonomous cars give continuous feedback to the driver because s/he will certainly be aware of the car driving by itself.

If the system even allows the choice of new exposure conditions (more extreme driving situations, higher speeds, etc.), the likelihood RA increases as well. Furthermore, RA is to be expected if subjective safety is increased by the system with the driver being convinced of her/his improved capability to handle critical situations. Finally, depending on personality, motivation, driving style, or current mood of the driver, the potential provided by the system could be (mis-)used in order to drive more experimentally or fun-oriented. The driver could "just for fun" try out maneuvers s/he wouldn't dare without the assistance of the system or use the system as a support for sensation seeking.

Gründl (2005) applies the scheme to a number of driver assistance systems in order to rate their impact on safety. The author analyzed more than 300 accidents by interviewing the drivers and by technically reconstructing the accident details. The *automatic emergency brake*, he concludes, is only marginally prone to RA effects because it is virtually unperceivable by the driver aside from (near) crash situations (criteria 1) and it does not play any role in regular driving (criteria 2--5). On the contrary, with *night vision* and *adaptive curve light*, the RA likelihood is rated particularly high. According to the author, the effects of the latter systems are permanently perceivable (criteria 1, 2). Furthermore, the enhanced vision leads to an increased scope of action (criteria 3) as drivers, who used to avoid driving at night, might now become less reluctant; others might also drive faster (also criteria 4,5). According to (Gründl, 2005) fully autonomous cars are considered critical by most experts, because the driver is absolved from responsibility and liabilities after a crash are unclear.

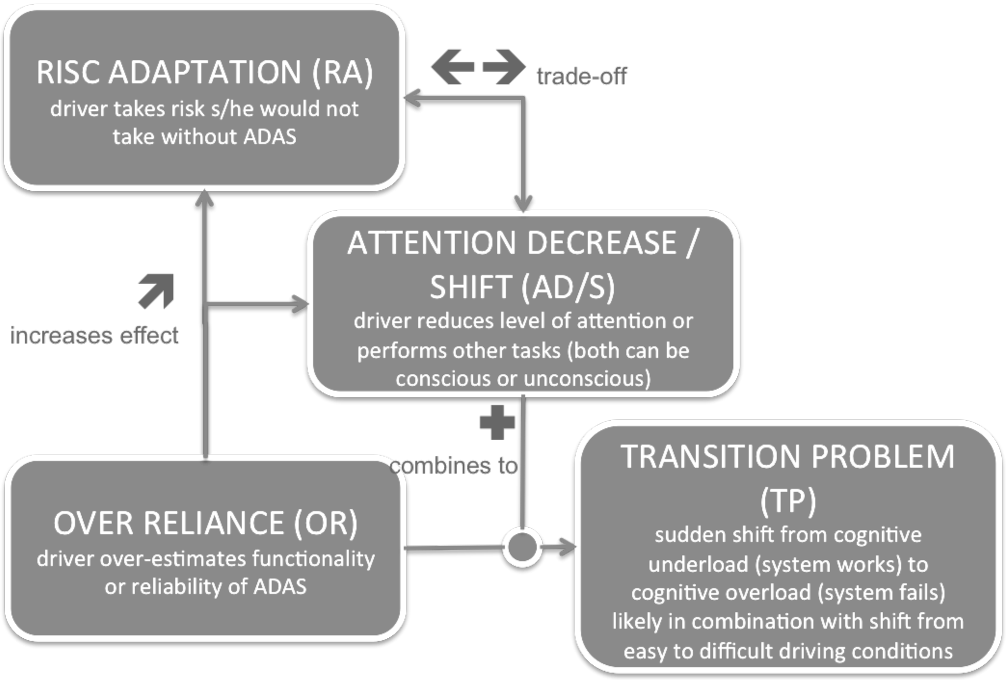
RISC ADAPTATION (RA)
driver takes risk s/he would not take without ADAS
trade-off
ATTENTION DECREASE / SHIFT (AD/S)
driver reduces level of attention or performs other tasks (both can be conscious or unconscious)
increases effect
combines to
TRANSITION PROBLEM (TP)
sudden shift from cognitive underload (system works) to cognitive overload (system fails) likely in combination with shift from easy to difficult driving conditions
OVER RELIANCE (OR)
driver over-estimates functionality or reliability of ADAS

Figure 1. Summary of various possible negative effects ADAS may have on safety

Figure 1 summarizes the various possible negative effects, which ADAS may have on safety. The nodes of the diagram are mainly based on Gründl (2005) as described above and the edges represent mutual influences between them. There is likely to be a trade-off between risk adaptation RA and attention decrease or shift (AD/S) in a sense that if the driver takes higher risks s/he will not be simultaneously less attentive and vice versa. This appears to be plausible, as taking a higher risk normally rises people's the stress level and physical arousal. On the other hand somebody who is distracted from driving or who is getting tired will rather compensate the attention decrease/shift for example by decreased speed or larger distance to a lead vehicle. As a result the driver has more time for reactions to an event in the environment. However, over-estimating the functionality or reliability of the system (OR) might increase the effects of both RA as well das AD/S. That means that the sheer trust in the system, even if its not fully justified, might make us drive riskier or become less attentive to the primary driving task.

In extreme cases, the transition problem (TP) can be intensified by a combination of OR and AD/S: after a long drive on a relatively empty freeway supported by a lane keeping assistant or possibly even by a fully autonomous car, the driver's attention has sunk to a very low level. All of a sudden, s/he approaches an unclear and confusing location – and just in that moment manual override becomes necessary because the system fails. Thus, over reliance is an important factor in the overall equation that results in negative effects on safety.

Our basic assumptions are supported by signal detection theory, which describes the relation between automation false alarms (FA)/ misses (MISS) and operator dependence (Meyer, 2001). It has been demonstrated, that an increase in automation false alarms decreases the operator's *compliance* resulting in longer response time to alerts and in some cases operators would even disregard those alerts (Dixon et al., 2007). An increasing automation's miss rate, on the other hand side, 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.

BEHAVIORAL EFFECTS OF LOCAL DANGER ALERTS

Local danger alert (LDA) is an important function of ADAS to improve the safety of driving. Besides directly sensing the environment to detect danger (Gavrila, 2001), recent advances in inter-vehicle communication technology (e.g. wireless ad-hoc networks car-2-car communication) further allow the exchange of information between cars (Kosch, 2004). This enables a much wider application of local danger warnings, as drivers can be alerted to approaching danger that is not yet visible. We focus on a scenario where drivers are warned about road obstacles that are a short distance ahead but not yet visible (e.g. due to a bend of the road or a leading vehicle), therefore requiring an immediate reaction (Mahr et al. 2010, Cao et al. 2010).

Table 1. Analysis of local danger warnings according to the criteria illustrated in Figure 1

Nr.	Criteria	In the experiment	In real life	Autonomous Cars
1.	interaction with the system	+ frequently	○ sparsely (obstacles appear not very often in real life)	○ depends on the specific characteristics of the system.
2.	immediate feedback	+	+	+
3.	widening of action scope	- (action scope is limited by experimental task	+ (knowing that there is no obstacle behind that bend might allow higher driving speeds)	∘ depends on the specific characteristics of the system
4.	increasing subjective safety	○ (is to be expected but there is certainly a different notion of ”safety” in an laboratory situation than in real life)	+	+
5.	superimposes ’acting-out’ tendencies	- (acting out does not happen in the experiment)	+ (corresponds here to widening of action scope)	○ (corresponds here to widening of action scope).

Table 1 analyses this particular type of local danger warnings according to the criteria introduced by (Pfafferoth & Huguenin, 1991) with respect to the schema illustrated in Figure 1. Generally, there is a difference between the experimental situation and real life. First of all, obstacles appear much more frequent during the experiment that on the road and thus the interaction with the system happens more frequently. Note, however, that an LDA system is likely to take effect at least once or twice on a long journey. Therefore, the driver is able to gain experience with the system and s/he will certainly be aware that it exists (as opposed to the automatic emergency brake example described above). Nevertheless, due to the higher value with respect to this criterion, it is to be expected that (yet to be described) effects are stronger in the experiment than they are in real life. On the other hand side, with some of the criteria the relation is vice versa: The experimental situation allows no widening of action scope as it is limited by the task and there is most probably no superimposing of risky behavioral tendencies as the experiment generates very little fun. Also, the increase in subjective safety might have a less strong effect because the lives of the subjects are not endangered. Altogether, it can be argued that diminishing and amplifying factors counterbalance each other thereabout.

Table 1 also anticipates the relation between ADAS types investigated here and fully autonomous cars at the far end of the spectrum. Whether or not the driver interacts frequently with the system depends: of course, sparse interaction is the basic idea of autonomous systems. On the other hand side, it is likely that in large-scale field test, which will become necessary in order to prepare dissemination of the technology, parameter-setting or other kind of interaction will be rather frequent. As discussed earlier, the feedback will in any case be very direct. With more elaborate technology, it is even imaginable that widening of action scope respectively enhancement of risky driving take place, for example if the technology allow driving faster or cope with more difficult terrain that the (particular) human driver.

CONCLUSION

ADAS technology is susceptible to behavioral impacts such as attention decrease/shift and over-reliance. A system's reliability and sensitivity can be viewed in terms of false alarm rate and miss rate. By varying the threshold settings for the decision criterion, designers are often able to trade one against the other (Parasuraman et al., 1997), as previous research has stated that false alarms could be more harmful to a user's performance than miss rate (Bliss, 2003; Dixon et al., 2006; Wickens & Dixon, 2007), while on the other hand a certain level of miss rate is tolerable, especially in demanding multi-tasking situations. Therefore the decision criterion for warnings should be set conservatively so drivers are aware of the possible (but relatively rare) misses and will therefore stay vigilant on the road. Of course the base rate of the critical situations (in which the system should become active) also needs to be considered for every single ADAS, as this particularly influences the false alarm and miss rate and therefore the perceived reliability. For the local danger warnings the base rate is probably higher than for systems like the emergency brake, but clearly lower than for other ADAS like the high beam assistant. As mentioned earlier a local danger alert is still likely to occur about one or two times during a long trip.

It was suggested that system reliability levels of 70% to 75% represent an optimum threshold for imperfect reliability assistance (Wickens & Dixon, 2007). Accordingly, a 75% reliable system would be better for the use on the road than one with 99% reliability. No technical system will be 100% failure-free – especially when taking into account driver „failures" like deactivating the system accidentally or not perceiving or realizing the system output at all. For example if a driver does not see, hear or feel a warning this seems to be a system failure for her/him. Even a technically perfect system would most probably result in a critical miss rate that is only close to 0%, leading to very rare misses which are particularly dangerous. Our research (Mahr et al., 2010) veers towards this argumentation, as rare misses potentially lead to severe performance decline accompanied by rising stress levels. This arises questions on the rollout strategy for fully autonomous cars, either into large-scale field test or into practice. The technology is susceptible for behavioral impacts such as attention decrease/shift and transition problems in combination with over-reliance. At the same time, autonomous cars will unlikely be 100% failure free. To sum up, the behavioral effects have to be taken into account, when setting the decision criterion of any ADAS and engineers must ensure that drivers are always aware that the system may fail.

REFERENCES

Bliss, J. (2003). An Investigation of Alarm Related Accidents and Incidents in Aviation. *Intern. Journal of Aviation Psychology*, *13*, 249–268.

Brookhuis, K. A., de Waard, D., & Janssen, W. H. (2001). Behavioural impacts of Advanced Driver Assistance Systems – an Overview. *European J. on Transportation and Infrastructure Research*, *1*(3), 245–253.

Cao, Y., Mahr, A., Castronovo, S., Theune, M., Stahl, C., & Müller, C. (2010). Local Danger Warnings for Drivers: The Effect of Modality and Level of Assistance on Driver Reaction. *Proc. International Conference on Intelligent User Interfaces (IUI) 2010*. Hong Kong.

Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2006). How Do Automation False Alarms and Misses Affect Operator Compliance and Reliance AEROSPACE SYSTEMS: Cognitive Factors in Aviation. *Proc. Human Factors and Ergonomics Society 50th Annual Meeting,* 25–29.

Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007). On the Independence of Compliance and Reliance: Are Automation False Alarms Worse Than Misses? *Human Factors*, *49*(4), 564–572.

Gavrila, D. M. (2001). Sensor-Based Pedestrian Protection. *IEEE Intelligent Systems*, *16(6)*, 77-81.

Gründl, M. (2005). *Fehler und Fehlverhalten als Ursache von Verkehrsunfällen*. University of Regensburg.

Kosch, T. (2004). Local danger warning based on vehicle ad-hoc networks: Prototype and simulation. *1st Intern. Workshop on Intelligent Transportation (WIT 2004)*.

Mahr, A., Cao, Y., Theune, M., Schwartz, T., & Müller, C. (2010): What if it Suddenly Fails? Behavioural Aspects of Advanced Driver Assistant Systems on the Example of Local Danger Alerts. *Proceedings of 19th European Conference on Artificial Intelligence (ECAI 2010)*, 1051-1052.

Meyer, J. (2001). Effects of Warning Validity and Proximity on Responses to Warnings. *Human Factors*, *42*, 563–572.

Parasuraman, R., Hancock, P. A., & Olofinboba, O. (1997). Alarm Effectiveness in Driver-centred Collision-warning Systems. *Ergonomics*, *40*(3), 390–399.

Pfafferoth, I., & Huguenin, D. (1991). Adaptation nach Einführung von Sicherheitsmassnahmen – Ergebnisse und Schlussfolgerungen aus einer OECD-Studie. *Zeitschrift für Verkehrssicherheit*, *73*(1), 71–83.

Wickens, C. D., & Dixon, S. R. (2007). The Benefits of Imperfect Diagnostic Automation: A Synthesis of the Literature. *Theoretical Issues in Ergonomics Science*, *8*(3), 201–212.