Advances in Transportation, Vol. 60, 2022, 33–39
https://doi.org/10.54941/ahfe1002431

Confidence Horizon for a Dynamic Balance Between Drivers and Vehicle Automation: First Sketch and Application

Nicolas D. Herzberger[1], Marcel Usai[1], and Frank Flemisch[1,2]

[1]RWTH Aachen University, Institute of Industrial Engineering and Ergonomics (IAW), Eilfschornsteinstr. 18, 52062 Aachen, Germany
[2]Fraunhofer FKIE, Fraunhoferstr. 20, 53343 Wachtberg, Germany

ABSTRACT

Advances in operator state monitoring and enhanced capabilities of automated vehicle systems will enable conditionally and highly automated vehicle systems (SAE level 3 and 4) in the near future. Possible transitions of control between driver and automation including handovers of the dynamic driving task to the driver will pose particular challenges. Several recent accidents show that understanding, exploring, designing, and testing these complex socio-technical systems not only in normal conditions, but especially at system limits and system failures is not only of scientific interest, but a matter of life and death. In the Confidence Horizon concept, the capabilities of the driver are continuously compared with those of the automated subsystem, resulting in two horizons: First, the technical subsystem's confidence in its own ability to safely control the vehicle, and second, the technical subsystem's confidence in the driver's ability to take over control. This allows to quickly identify whether transitions between different levels of automation are safe, whether a balanced control distribution is given, or whether, when and how a minimum risk maneuver may be necessary. The concept can thus serve as a first approach to describe, visualize and implement possible cooperation between users and automated vehicles. In this way, future highly and cooperatively automated vehicles could be improved by revealing safety-critical transitions between driver and vehicle at an early stage. This paper first describes the confidence horizon concept and then presents a first implementation.

Keywords: Automated driving, Human systems integration, Cooperative systems interaction, Cooperative automation, Highly automated systems

INTRODUCTION

"May you live in interesting times": This Chinese wish, sometimes interpreted as curse, also applies to today's times of the early 20s of the 21st century: Not only are a Covid-19 pandemic, global warming and military threats overshadowing our lives, but long term scientific and technical revolutions are once again coming to a point where they could fundamentally change the way we work and live. One example is virtualization as a more detached way of interacting with others and the world, e.g. in developments such as

AR, VR, XR and the Metaverse. Another interesting development is that of the increasing cognitive abilities of machines, often referred to as automation and artificial intelligence (AI). While some engineers and scientists dream of autonomous machines, e.g., autonomous driving, autonomous factories, or autonomous weapon systems, others call this a hype that might collapse soon (e.g., as described in the Gartner hype cycle) or a threat to society and humans (e.g., as described in Nick Bostrom's 'Superintelligence', 2017).

In this long term tension field between human and automation/AI, between fully manual or fully autonomous, an increasing part of the scientific and technical community is exploring a middle ground, integrating humans, machines, organizations and the environment in a way that they form well balanced systems. For an overview of this art, craft and science of balanced Human Systems Integration see e.g. Flemisch et al. (2021). Concepts such as human machine symbiosis (e.g. Licklider 1960ff), human machine cooperation, cooperative automation (Flemisch et al. 2016a), or shared and cooperative control (e.g. Abbink et al., 2012) describe that this combination of humans and machines is more than just a sum, that they form joint cognitive and cooperative systems that are more capable than any of the subsystems alone. An essential part of this cooperation is that conflicts, e.g. when resources are limited or within transitions of control, are detected, arbitrated and mediated in a sufficient way (e.g. Baltzer et al. 2018; Flemisch et al. 2020).

COOPERATIVE DRIVING

Applied to intelligent transportation systems, the system design described in this paper follows a general approach of human-machine cooperation as described in Flemisch et al. 2016b and modified by Flemisch et al. 2019 and Usai et al 2020: For a given cooperative task, the cooperation is divided into four main layers of cooperation: A strategic (e.g. navigational), tactical (e.g. maneuver and trajectory planning) and an operational (e.g. stabilization of vehicle on planned trajectory, lane keeping etc.) layer, as well as a cooperational layer, on which human and machine communicate about the topic of cooperation itself and how their cooperation is designed. On each layer, human and machine work on a joint action. This general framework is depicted in Figure 1. With the help of a mediator, human and machine can share and transfer control on each layer to match the driver's abilities and intentions. To provide more transparency and an easier understanding, the automation's behavior follows a model used to describe human information processing by Parasuraman et al. 2000. In addition, the mediator observes the human to predict its intention and involvement in a given task to assess whether a control transfer from the machine to the human would be possible in a given situation.

CONFIDENCE HORIZON CONCEPT

Even in the context of cooperative driving, the driver will still have to act as a fallback option in level 3 or take over the driving task again after leaving the ODD in level 4 (SAE international, 2021). To ensure the safety of the

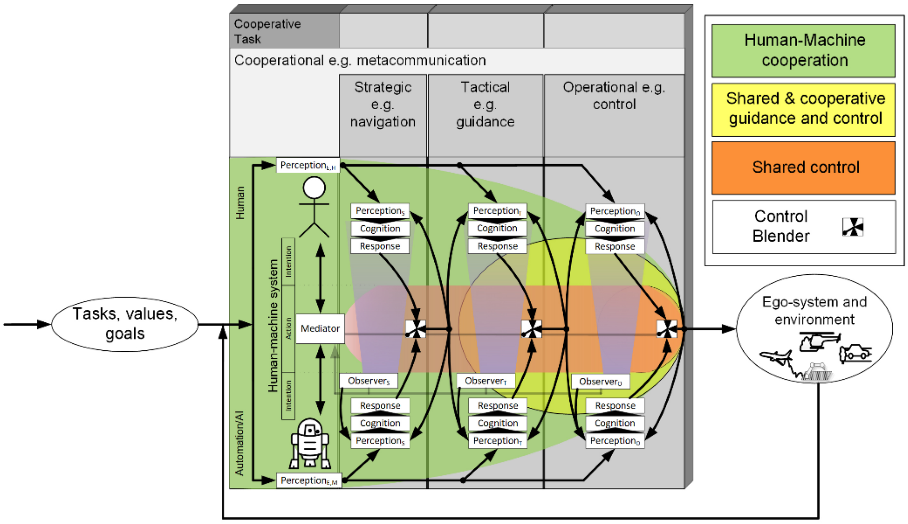

Figure 1: Framework for a mediated human-machine cooperation, in which the mediator distributes authority to control based on the confidence in the measured abilities of human and automation to be able to execute control of the motion control task, or, as displayed, a cooperative task in general. (Framework based on Flemisch et al. 2019; Usai et al. 2020.)

system in these cases, it is crucial to identify potential safety-critical transitions (Flemisch et al. 2019). However, to identify these, the transition process first needs to be understood. When the driving task is handed over to the driver, the control deficit due to the lack of driver involvement is the central criterion for identifying potential risks (Herzberger et al. 2020). Figure 2 shows a model which can be used to identify these potential safety critical transitions. Depicted are two possible transitions from highly automated to lower levels of automation: On the one hand, the green arrow and ovals represent a safe transition from highly automated to partially automated driving. Here, the blue (system-side task completion) and orange areas (driver takeover capability) overlap, so that no control deficit is to be expected. On the other hand, a safety-critical transition from highly to manual driving is shown (red arrow and oval). Here, the blue and orange areas do not overlap, indicating that the added capabilities of the driver and the automated driving system together are not sufficient to execute the driving task safely. This critical deficit of control is also apparent by the white triangle between the respective capabilities.

However, the control distribution model is only a first step on the way to implementing cooperative automated driving systems. In the future, the assessment of human capabilities as well as an evaluation of the capacities of the technical system could enable confidence horizons for ensuring safe transitions. This confidence horizon concept will be explained in the following using an example use-case from the priority program *Cooperatively Interacting Automobiles* of the German Research Foundation (SPP 1835). Here, the transition of the driving task from automated system to human driver is examined in the context of a critical situation on the highway (see Figure 3,

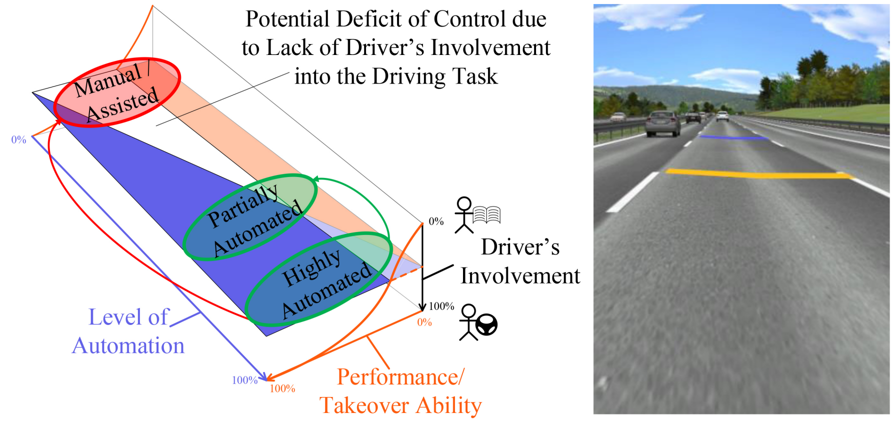

Figure 2: Potential deficit of control (Adapted from Flemisch et al. 2019), theoretical (left) and first applied (right) model.

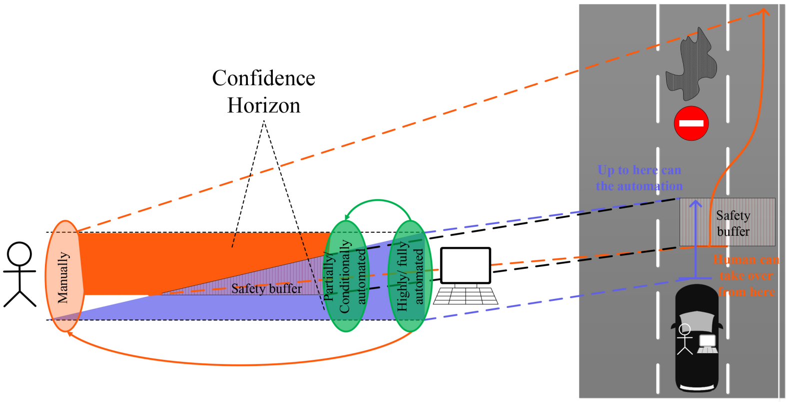

Figure 3: Confidence horizon concept with an example of a potential safety buffer.

right). In this situation, the confidence horizon concept estimates the distance from or up to which the human is able to perform the driving task safely. These points are compared with those from or up to which the automated driving function is capable of safely performing the driving task, with the aim of identifying and counteracting possible risks during takeovers or handovers.

Figure 3 (left) illustrates in orange the capabilities of the human and in blue those of the automated driving system. The overlap of these areas represents a safety margin. In this exemplary use-case (right), a transition from highly to partially automated driving is possible because there is an overlap of capabilities (safety buffer), visualized by the bars in front of the ego vehicle "Human can take over from here" and "Up to here can the automation". A transition to manual driving, however, would not be possible in this example, since there is no overlap of the capabilities in this area, e.g., due to a lack of human involvement, which is why a transition could lead to a potentially critical control deficit.

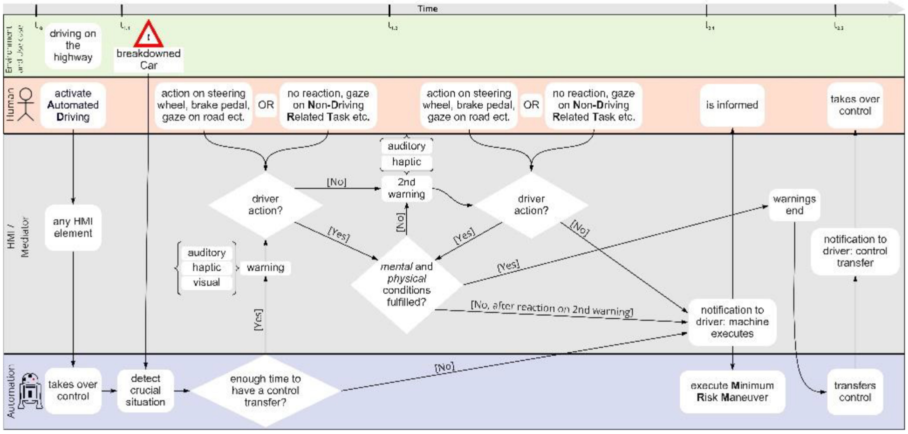

Figure 4: Flow chart diagram explaining the interaction between human and automation in the use-case described above.

In order to transfer this theoretical model to real-world applications, a first application of the confidence horizon concept is presented in the following.

FIRST APPLICATION

For the above described use-case, Figure *4* shows a flow chart of the designed human-machine interaction in that case. The design combines an inform-warn-intervene with collision avoidance pattern. At time t_0, the driver has activated the automation system and might have focused on another task. Neither following nor executing the driving task, it has to be assumed, that no situation awareness is present. The potential system failure begins at time $t_{1.1}$, when the automation system detects a breakdown vehicle or another obstacle ahead while driving in the center lane. Before triggering a take-over request (TOR), it evaluates, whether there is enough time remaining to have a control transfer (calculating the orange bar in Figure 3) and if yes, triggers a two-stage multimodal warning (see Zhang et al. 2021 or Guo et al. 2021, i.e. escalation steps) to get the driver involved.

To help building situation awareness of the driver as quick as possible, information on the environment and the human and automation confidence horizon bars are displayed in the contact analogue Head Up Display (HUD). In this scenario, this translates into two possibilities: either the human takes over early enough to solve the situation, or the automation has to use up its time reserve to perform a, possibly still risky or non-optimal, minimum risk maneuver (MRM) (Hoeger et al., 2011). This happens, in case there is not enough time for a hand over from the beginning, or the human does not want to take over or the automation calculates, that there is a high chance, that the human will not be able to build enough situation awareness to take over in time (Herzberger et al. 2018). In that case, the human is notified that the automation prohibits a take-over due to a MRM.

OUTLOOK

This paper can only give a small insight in the design of intuitive human-machine cooperation in the domain of automated driving. However, it already provides fundamental concepts aimed at improving the efficiency of human-machine cooperation, especially in dangerous and time-critical situations, without losing intuitiveness or requiring comprehensive training. Its target use-cases might seem rare, but due to the criticality are even more important ones.

In a next step, the authors evaluate the implementation of the confidence horizon design in a driving simulator study against a conventional setup in system boundary and failure use-cases.

ACKNOWLEDGMENT

This publication was funded within the Priority Programme 1835 "Cooperative Interacting Automobiles (CoInCar)" of the German Science Foundation (DFG).

REFERENCES

Abbink, D. A., Mulder, M., & Boer, E. R. (2012). Haptic shared control: smoothly shifting control authority?. *Cognition, Technology & Work*, *14*(1), 19–28.

Baltzer, M.; Flemisch, F. (2018): Arbitration: When Time for Interaction Mediation in a Human Machine System is running out. Ed. Flemisch, F.; Abbink, D., Cognition, Technology & Work, Special Issue "Shared and cooperative control of safety critical systems".

Bostrom, N. (2017). *Superintelligence*. Dunod.

Flemisch, F.; Winner, H.; Bruder, R. & Bengler, K. (2016a). Cooperative Vehicle Guidance & Control. In. Winner, H.; Hakuli, S.; Lotz, F. & Singer, C. *(Eds.)*: *Handbook of Driver Assistance Systems,* Wiesbaden: Springer Nature.

Flemisch, F., Abbink, D., Itoh, M., Pacaux-Lemoine, M. P., & Weßel, G. (2016b). Shared control is the sharp end of cooperation: Towards a common framework of joint action, shared control and human machine cooperation. *IFAC-PapersOnLine*, *49*(19), 72–77.

Flemisch, F., Abbink, D. A., Itoh, M., Pacaux-Lemoine, M. P., & Weßel, G. (2019). Joining the blunt and the pointy end of the spear: towards a common framework of joint action, human–machine cooperation, cooperative guidance and control, shared, traded and supervisory control. *Cognition, Technology & Work*, *21*(4), 555–568.

Flemisch, F. O., Schwalm, M., Meyer, R., Altendorf, E., Lennartz, T., Schreck, C., Bavendiek, J. & Herzberger, N. D. (2019). Human system integration at system limits and system failure of cooperatively interacting automobiles: concept and first results. *IFAC-PapersOnLine*, *52*(8), 93–98.

Flemisch, F.; M.-P. Pacaux-Lemoine, F. Vanderhaegen, M. Itoh, Y. Saito, N. D. Herzberger, J. Wasser, E. Grislin, M. Baltzer (2020): Conflicts in Human-Machine Systems as an Intersection of Bio- and Technosphere: AiAIDo: IC-HMS Conference on Human Maschine Systems.

Flemisch F. (2021). Balanced Human Systems Integration. Keynote at IHSI 2021.

Guo, H., Zhang, Y., Cai, S., & Chen, X. (2021). Effects of Level 3 Automated Vehicle Drivers' Fatigue on Their Take-Over Behaviour: A Literature Review. *Journal of Advanced Transportation*, *2021*.

Herzberger, N. D., Eckstein, L., & Schwalm, M. (2018). Detection of Missing Takeover Capability by the Orientation Reaction to a Takeover Request. 27. In Aachen Colloquium Automobile and Engine Technology, Aachen (pp. 1231–1240).

Herzberger, N. D., Meyer, N. R. & Flemisch, F. O. (2020). Systemergonomie für kooperativ interagierende Automobile: Nachvollziehbarkeit des Automatisonsverhaltens und Eingriffsmöglichkeiten des Menschen. *IAW Spectrum*, 15(*1*), 8–9.

Hoeger, R.; Zeng, H.; Hoess, A.; Kranz, T.; Boverie, S.; Strauss, M.; Jakobsson, E.; Beutner, A.; Bartels, A.; To, T.-B.; Stratil, H.; Fürstenberg, K.; Ahlers, F.; Frey, E.; Schieben, A.; Mosebach, H.; Flemisch, F. O.; Dufaux, A.; Manetti, D.; Amditis, A.; Mantzouranis, I.; Lepke, H.; Szalay, Z.; Szabo, B.; Luithardt, P.; Gutknecht, M.; Schömig, N.; Kaussner, A.; Nashashibi, F.; Resende, P.; Vanholme, B.; Glaser, S.; Allemann, P.; Seglö, F. & Nilsson, A. (2011) HAVEit Deliverable D61.1, Final Report.

Licklider, J. C. (1960). Man-computer symbiosis. *IRE transactions on human factors in electronics*, (1), 4–11.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, *30*(3), 286–297.

SAE International (2021). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (No. J3016_202104)*. 400 Commonwealth Drive, Warrendale, PA, United States. SAE International.

Usai, M., Meyer, R., Nagahara, H., Takeda, Y., & Flemisch, F. (2020). Towards a Truly Cooperative Guidance and Control: Generic Architecture for Intuitive Human-Machine Cooperation. In *International Conference on Applied Human Factors and Ergonomics* (pp. 92–98). Springer, Cham.

Zhang, W., Zeng, Y., Yang, Z., Kang, C., Wu, C., Shi, J., ... & Li, H. (2021). Optimal Time Intervals in Two-Stage Takeover Warning Systems With Insight Into the Drivers' Neuroticism Personality. *Frontiers in Psychology*, 157.