The Notion of Controllability in an autonmous vehicle context

Helen Monkhouse, Ibrahim Habli, John Mcdermid

The Notion of Controllability

in an autonmous vehicle context

Helen Monkhouse[†‡], Ibrahim Habli[†], John McDermid[†],
University of York[†]
Protean Electric Ltd.[‡]

***Abstract*—For decades, the automotive industry has used the notion of controllability as a way of assessing the likelihood that, when subject to a hazardous situation, the vehicle driver avoids an accident. The introduction of semi- and fully-autonomous vehicle systems has the potential to fundamentally alter the driver's tasks and role during all or part of the journey, and challenges the notion of controllability as used in existing automotive risk models. This paper explores the notion of controllability and highlights areas where increasing autonomy may impact the use of controllability in risk assessment.**

Keywords—controllability, system safety, ADAS, autonomous

I. INTRODUCTION

When developing a new automotive feature, or adding an existing feature onto a new vehicle platform, consideration must be given to potential hazards. When hazard risk is determined using the ISO 26262 standard [1], the severity of the accident, the probability of exposure to, and the controllability of, the hazardous events are used to determine the Automotive Safety Integrity Level (ASIL). The activities undertaken to develop and integrate the feature can then be tailored for the highest ASIL identified; with the highest integrity level demanding the highest design rigour [2].

When the automotive industry considers controllability, the rating for a hazardous event often relies on domain experts combining their knowledge of the feature behavior with their experience in vehicle handling. This may seem un-scientific. However, this is a mature industry that has slowly evolved its products over decades, and the user interface has remained relatively stable for a century, thus a subjective engineering judgement based assessment seems defensible.

However, today this gradual product evolution is being challenged with the introduction of autonomous vehicle features. This not only changes the driving task and the way in which the driver interacts with the vehicle, but also challenges the implicit assumptions made about the driver's ability to control the vehicle should a failure occur. This, in turns, leads us to question the validity of current automotive risk models, particularly those used in standards such as ISO 26262 [1].

This paper is organised as follows. Section II discusses the origins of controllability and the assumptions made about the driver, their role in the control loop and the driving task. Familiar vehicle systems are used to illustrate the notion of controllability when the driver is viewed as part of the control loop for a fully manual vehicle. Section III discusses two pivotal human factors papers and explores how greater autonomy may affect the notion of controllability. Emerging Advanced Driver Assistance Systems (ADAS) are used to illustrate the points raised. Finally the paper concludes with a summary and a discussion of next steps in Section IV.

II. CONTROLLABILITY

A. Origins of Controllability

In control theory controllability is viewed as the ability to 'steer' a dynamic system from an initial to a final state using admissible inputs [3]. The notion traces back to the work of Kalman in the 1960s [4].

From a vehicle handling perspective controllability refers to the relationship between the driver's ability and the vehicle's handling qualities [5]. In the aerospace domain various studies have been undertaken to facilitate the qualitative evaluation of controllability. One notable pilot-rating scheme is the Cooper Harper Rating Scale devised in the late 1960s to help pilots and engineers evaluate aircraft handling and stability [6].

The notion of controllability can also be found in civil and military aerospace standards. MIL-STD-882C [7] included controllability in its software control categories, with Category I software exercising full autonomous control over the system (with no potential for intervention to mitigate the hazard), Category IIb software providing display information needing immediate operator action, and so on. Civil aviation standards use a similar notion of controllability, for example DO-178C [8] includes system *failure categories* that consider "the flight crew's ability to cope", from which software levels are derived.

From an automotive system safety perspective the notion of controllability traces back to the Dedicated Road Infrastructure for Vehicle Safety in Europe (DRIVE) I project, DRIVE Safely, undertaken in the early 1990s [9]. Here controllability is the probabilistic attribute linking a hazardous event to an accident; by indicating how likely the driver will be able to control the hazardous situation and thus avoid harm. Thus a hazard categorised as *nuisance only* implies a zero probability of that event becoming an accident. Whereas, a hazard categorised as *uncontrollable* cannot be influenced by human intervention to give a positive outcome.

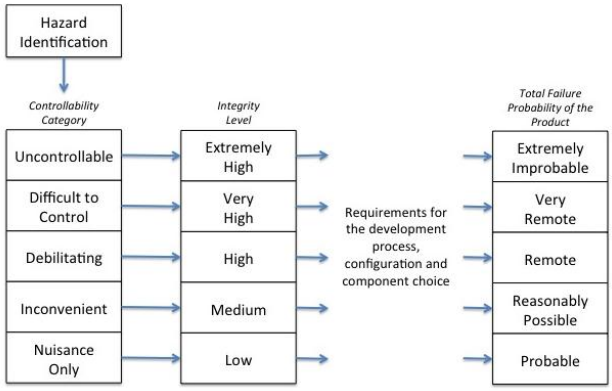

Fig. 1 Controllability and Integrity Categories from DRIVE

The DRIVE Safely project categorised controllability into five levels (Fig. 1) linking it to integrity levels directly. With the driving environment complicating hazard classification - at the extreme it is possible to identify numerous, even fatal, outcomes from the same seemingly benign hazard [10] - controllability categorisation was left to expert engineering judgement.

The 1994 MISRA Guidelines [11] included example 'influencing factors' to help guide controllability categorisation. This was refined further by the 2007 Guidelines [12]; with the risk model separating controllability and severity and Appendix D including a method for considering and using 'influencing factors' under various circumstances (including greater autonomy). 'Influencing factors' were not included in ISO 26262 [1], with a less structured assessment being required.

The Controllability of Automotive Safety Targets (CAST) project [13] used vehicle simulation to validate the assumptions made by the DRIVE Safely project about the inverse relationship that exists between controllability and the probability of having an accident following a hazard. The study was able to correlate the simulation results with an expert assessment of controllability completed beforehand, showing the soundness of the concept.

B. *Assumptions Made*

Although not explicitly stated, ISO 26262 makes three assumptions about the driver. Firstly, the driver is always part of the control loop; with the origins of this assumption routed in the 'Vienna Convention' [14]. Secondly, the driver is integral to control, as illustrated by Fig. 2 [12], implying that the driver is fully aware of their surroundings. Thirdly, that the 'safe state' can be to shut-down the system; i.e. 'fail passive'.

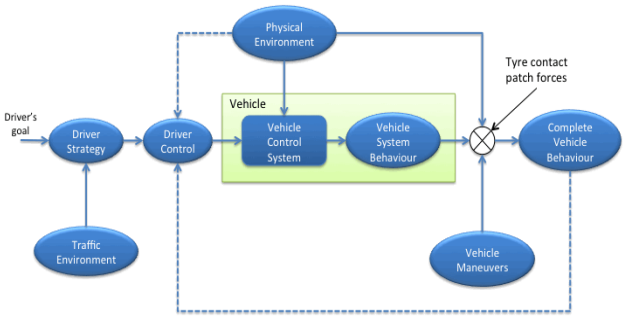

Fig. 2 Control System View of the Vehicle

C. *Controllability of Familiar Automotive Examples*

To illustrate the notion of controllability the following vehicle features are discussed: an engine control system, an air suspension system, and an electric driveline incorporating in-wheel motors (see Fig. 2).

Emissions legislation, fuel economy, performance and cost all contributed to engine control systems evolving from a purely mechanical design to a complex programmable control system. Although the earliest mechanical systems could cause *Unintended Acceleration* (for example, by the accelerator cable becoming stuck) the introduction of electronic throttle control systems increased focus on the hazard.

The early electronic control systems typically mimicked previous mechanical designs, whereas modern engine control systems utilise complex control algorithms expressed in the torque domain. This leads to a complex relationship between the acceleration pedal position and the resultant vehicle acceleration. Typically layered monitoring strategies [15] are used to continuously monitor the relationship between the driver's torque demand (via the accelerator pedal) and the delivered engine torque (estimated from measured engine parameters); any hazardous discrepancy between demanded and delivered torque is mitigated by limiting the engine's performance or ultimately by shutting it down.

This concept naturally reflects the notion of controllability and the assumptions described in Section II.B. For an experienced driver the longitudinal control task is generally a subconscious process, with the driver modulating the accelerator pedal position (*Driver Control*), based on their current *Driver Strategy* and given their knowledge of the current *Traffic Environment*, to achieve the desired vehicle acceleration. If vehicle acceleration (*Complete Vehicle Behaviour*) is too slow then the accelerator pedal can be pressed harder. Conversely, if acceleration is too great then the accelerator pedal can be released, and once completely released the brakes applied.

To mitigate the hazardous effects of engine control system failures, controllable acceleration targets can be defined, e.g. "*Vehicle positive longitudinal acceleration shall not exceed driver demand by > 1.5 m s^2 for longer than 1 s"*, knowing the driver's likely response [16]. The mitigating action "*outputs are electronically 'limited' to a fixed value*" can then be defined, thus ensuring that any engine torque anomalies detected are prevented from exceeding a level where the resultant acceleration might be difficult to control. With these targets defined the torque monitor's internal parameters can then be set to achieve the required response.

Controllability works well for air suspension systems of the type used on large luxury four-wheel drive sport utility vehicles to change the vehicle's ride height; with ride height being raised to facilitate off-road driving and lowered to make ingress and loading easier [17]. As selecting an incorrect ride height could affect the vehicle's handling characteristics (changing the relationship between *Driver Control* and *Complete Vehicle Behaviour* in Fig. 2) vehicle speed limits constrain when ride height changes can be made. However, a failure leading to the vehicle being at the wrong ride height, for a given vehicle speed, may result in a hazard.

Typically safety mechanisms detect failures within the air suspension system. Having detected a failure, and dependent on the current *Vehicle Manoeuvre*, the safety mechanisms can then either disable the system (*Fail Passive*) and issue driver warnings (dashboard warning lights, messages or audible chimes) or disable the system and limit the speed. In either case the failure mitigation makes assumptions about the driver's place within the overall control loop. In the first case, once the driver has been warned about the potential changes to vehicle handling they will change their *Driver Control* accordingly. In the second case, where the vehicle speed is

limited, the vehicle's operating envelope (or choice of *Vehicle Manoeuvres*) is constrained into the region where the vehicle is known to be controllable.

For novel technologies, such as an in-wheel motor electric driveline system [18] the notion of controllability is still valid because the assumption that the driver is a part of the control loop is still holds. Although the ability to control torque delivery to each road wheel independently is a vehicle dynamists dream, and a major benefit of in-wheel motors, applied incorrectly this torque asymmetry causes the hazard *Induced Yaw*. Empirical driver response studies [19] set yaw rate and lateral acceleration controllability limits giving a power asymmetry target of 30 kW / tonne [20]. Safety mechanisms monitor the torque being produced by a pair of motors, and either shut-down or modify motor control if the asymmetry exceeds the limit. Thus, system failures having the potential to adversely affect the *Vehicle Control System* can be detected and *Vehicle System Behaviour* influenced so that the *Complete Vehicle Behaviour* remains controllable by the driver.

III. The Impact of Greater Autonomy

To this point we have used *automation* and *autonomy* somewhat interchangeably. This has been done to facilitate a practical rather than theoretical discussion, as in the strictest sense all features discussed represent *automation*.

A. Pivotal Literature

Could Lisanne Bainbridge's seminal paper, "The Ironies of Automation" [22] provide insight into how the driver's relationship with the car will change with greater autonomy? And could it give insight into how engineering design assumptions may need to change to preserve safety?

Bainbridge discussed how greater automation in an industrial process context might actually increase rather than decrease human operator problems, particularly if automation changes the human's role to that of a supervisor required to intervene only under abnormal conditions. This relationship between the human operator and automation is discussed through a number of '*ironies*'; see below for examples.

A second paper potentially holding more clues is "Humans and Automation: Use, Misuse, Disuse, Abuse", written by Raja Parasuraman and Victor Riley 14 years after Bainbridge published her work. Again ideas presented by these authors are considered in the context of vehicle autonomy below.

B. The Use of Automation

As vehicle systems gain greater autonomy, it is perhaps logical to expect *Driver Control* to become a shared task, between the human driver and the autonomous system, but how does this affect controllability?

One motivation behind the introduction of autonomous vehicle systems is the reduction in accidents resulting from human error in the *Driver Control* task. The first challenge is choosing which tasks to automate, as deploying automation when it is not the right option is an abuse of automation, potentially leading to increased driver workload [21].

Automation has supported the driver in the longitudinal vehicle control task for many years. The earliest cruise control systems essentially enabled the driver to set a constant vehicle speed while manually undertaking the remainder of the *Driver Control* task. Assuming that the vehicle is in a *Physical Environment* where the driver has sufficient space and time in which to react, cruise control failures are normally easily controllable. This assumption can be met by constraining the use of cruise control, e.g. via a minimum set speed.

Adaptive Cruise Control (ACC) supports the *Driver Control* task further. Although the ACC fulfils the longitudinal control aspects of *Driver Control*, the driver is not relieved of all responsibility. In some operational situations, such as the preceding vehicle braking sharply, the ACC may require the driver to resume manual control because the ACC system does not have sufficient authority over vehicle brake pressure to achieve a full emergency stop. Therefore while ACC is active the driver inherits a new monitoring task, and the driver must understand ACC operation sufficiently to know when manual control may be required [21], and remain vigilant and ready to intervene when required [22].

A vehicle fitted with an ADAS, such as ACC, also places new responsibility on the driver, mainly knowing when to give longitudinal control to the ACC system and when to undertake *Driver Control* manually. Automation misuse can occur when users become over-reliant on automation [21]. For example attempting to use ACC when environmental factors (in fog or on very winding roads) limits system performance or leads to unexpected system behaviour and potentially system disuse because the driver's confidence in the system is lost [21].

People use different strategies when choosing whether to use automation [21] adding complexity and uncertainty into the design task. Also, gathering the verification evidence that enables the design team to demonstrate system safety becomes a far bigger task. This is because relatively simple automation leads Fig. 2 to become multi-dimensional; with a separate dimension existing for each *Driver Control* combination.

C. The Driver's Task with Autonomy

As previously stated one intuitively knows that the driver's role (Fig. 1) will change, but what might the new role become and will the driver remain effective in that new role?

ACC showed how the driver's role might become a monitoring task. The irony here is that humans are less good at monitoring automated tasks than they are at carrying out the task manually [22]. If the manual task was largely achieved subconsciously then automating that task and asking the driver to monitor it could actually add to the driver's workload [21]. For example, with Lane Keeping Assistance (LKA), automation supports a subconscious task. Steering is a *Driver Control* task requiring little driver cognition; the driver practices constantly during each journey, so it quickly becomes completely subconscious even for the most novice driver. LKA is designed to support lateral vehicle control by applying a corrective torque (either through the steering system or by an asymmetric brake application), if the driver allows the vehicle to stray from its lane.

The automation of a task of this type raises questions about the actions that should be taken in the presence of failure. For systems like LKA, *Fail Passive* is not a viable option. It also challenges the controllability assumptions that would have been made regarding other systems fitted to the vehicle. For example the power assisted steering system and the hazard loss of steering assist. The introduction of LKA effectively adds a use case. As such, will controllability assumptions made previously still hold true? Additionally, the use of LKA may lead to loss of situational awareness which is known to adversely affect driver reaction times [23], complicating further the choice of mitigation.

D. The Designer's Task with Autonomy

Both papers [21, 22] discuss the importance of the design team's view of the human operator. If ADAS features are developed with the sole motivation of removing the unreliable and inefficient driver from the control loop then Bainbridge warns of two potential outcomes [22]: design teams may target the easily automated *Driver Control* tasks first leaving the driver with the more difficult tasks to be performed manually!

An inevitable consequence of greater autonomy is also increased complexity and as the control systems view of a vehicle may become a multi-dimensional problem; with the potential for subtle interactions between the dimensions. The desire to remove human error may simply move the errors to the design task, as "one cannot remove the human error from a system simply by removing the human operator" [21]. Thus today's vehicle is a complex system of systems having the potential to exhibit both good and bad emergent properties.

IV. SUMMARY

Human factors models have been used to 'test' the notion of controllability and the control system view of a vehicle (*Fig. 2*). Although these models work when there is no automation, even for function specific automation (like ACC) deficiencies become apparent. The most notable omission is perhaps the lack of feedback from *Vehicle System Behaviour* to *Driver Control* or *Driver Strategy*.

The notion of controllability has been used to good effect in the past. However greater autonomy challenges the model, which should be re-evaluated and the term controllability better defined. The Control Systems View of a Vehicle (*Fig. 2*) is indeed useful, but it needs to change to consider situational awareness and the multi-dimensional nature of the problem.

REFERENCES

[1] "ISO 26262: Road Vehicles - Functional safety," 2011.

[2] I. Habli, R. Hawkins, and T. Kelly, "Software safety: relating software assurance and software integrity," *Journal of Critical Computer-Based Systems*, vol. 1, pp. 364-383, 2010.

[3] J. Klamka, "Controllability of dynamical systems. A survey," *Polish Academy of Sciences Bulletin: Technical Sciences* vol. 61, pp. 335, 2013.

[4] T. Berger, T. Reis, "Controllability of linear differential-algebraic systems—a survey," in *Surveys in Differential-Algebraic Equations I*, pp. 1-61, 2013.

[5] M. Abe, *Vehicle handling dynamics: theory and application*: Butterworth-Heinemann, 2015.

[6] G. E. Cooper, R. P. Harper Jr, "The use of pilot rating in the evaluation of aircraft handling qualities," DTIC, 1969.

[7] "MIL-STD-882C: System safety program requirements,", *US Department of Defense*, 1993.

[8] "DO-178C Software Considerations in Airborne Systems and Equipment Certification," FAA, 1992.

[9] P. H. Jesty, T. F. Buckley, M. M. West, "The development of safe advanced road transport telematic software," *μProcessors & μSystems*, vol. 17, pp. 37-46, 1993.

[10] P. H. Jesty, K. M. Hobley, R. Evans, I. Kendall, "Safety analysis of vehicle-based systems," in *Proceedings of the 8th Safety-critical Systems Symposium*, pp. 90-110, 2000.

[11] "*Development Guidelines for Vehicle Based Software*" MISRA, 1994.

[12] "*Guidelines for safety analysis of vehicle based programmable systems"* MISRA, 2007.

[13] "Use of Controllability for Classification of Automotive Vehicle Hazards," MISRA Tech Rep., 2007.

[14] "Vienna convention on road traffic," UN, 1968.

[15] "Standardized E-Gas Monitoring Concept for Gasoline and Diesel Engine Control Units," eGAS Working Group, 2013.

[16] J. Birch, R. Rivett, I. Habli, B. Bradshaw, J. Botham, D. Higham, *et al.*, "Safety cases and their role in ISO 26262 functional safety assessment," *Computer Safety, Reliability, and Security*, pp. 154-165, 2013.

[17] I. Habli, I. Ibarra, R. S. Rivett, and T. Kelly, "Model-based assurance for justifying automotive functional safety," SAE Technical Paper 0148-7191, 2010.

[18] M. Ellims, H. Monkhouse, and A. Lyon, "ISO 26262: Experience applying part 3 to an in-wheel electric motor," in *IET 6th Sys. Safety Conf.*, pp. 1-8, 2011.

[19] A. Neukum, E. Ufer, J. Paulig, and H. Kruger, "Controllability of superposition steering system failures," *Steering tech*, 2008.

[20] M. Ellims, H. Monkhouse, D. Harty, and T. Gade, "Using Vehicle Simulation to Investigate Controllability," *SAE Journal of Alternative Powertrains*, vol. 2, pp. 18-36, 2013.

[21] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," *Human Factors*, vol. 39, pp. 230-253, 1997.

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

[23] N. Merat and A. H. Jamson, "How do drivers behave in a highly automated car," in *Proceedings of the 5th International Driving Symposium on Human Factors*, pp. 514-521, 2009.