Manuscript version
DOI original paper: 10.1109/ICHMS49158.2020.9209509

Establishing a modified CREAM approach for reliability evaluation

Chao He and Dirk Söffker[1]

***Abstract*—The human factors area is one of the core field of industrial safety. It is deeply recognized that human factors are the main cause of accidents and breakdowns in various industries. Physiological mental states of human operators, like fatigue or vigilance, are crucial for the evaluation of human operators reliability. However, the shortages of measurements of physiological mental states limit their application. So human reliability analysis (HRA) is developed and many techniques have been developed. Among these techniques, cognitive reliability and error analysis method (CREAM) is often applied and discussed. Due to the application limits of common performance conditions (CPCs) in original CREAM approach, it is advised to generate a new list of CPCs for the application domain if CREAM approach is applied to other domains. In this contribution, a new approach defining a situated and dynamical human reliability measure is established. The approach is based on the well-known CREAM approach, which is modified with respect to the use in dynamical context. The new list of CPCs, their levels and corresponding effects on human reliability are determined. A new index, human performance reliability score (HPRS), is proposed for the assessment of human operator reliability. Finally, the applicability and correctness of the newly established approach is verified by data analysis.**

***Index Terms*—human operators, physiological mental states, human reliability analysis, CREAM, CPCs, situated driving context**

I. State-of-the-art in human factors

Reliability of human operators is one of the most important factors affecting industrial safety. With the development of technology, accidents in industrial areas have decreased significantly, and less problems are caused by hardware or software in industry, but more accidents are related to human factors. The study [1] reported that 50 to 70 % of the risk at nuclear power facilities is caused by human errors. The study [2] suggested that human errors are the major contributory factor for shipping accidents as about 80 % of marine casualties are caused, at least in part, by some forms of human errors. In aviation area, most would agree that somewhere between 60 and 80 % of aviation accidents are attributable to human error [3]. The national highway traffic safety administration (NHTSA) states that 94 % of traffic accidents are related to human factors [4]. Therefore, human factors in industrial accidents should be studied and evaluated in a proper way. Two main research directions are generated, physiological mental states estimation and human reliability analysis (HRA). Physiological mental states are mainly considered for single human operator, while HRA can be used in single human operator or group operation in context. Take the driving context as an example, in existing research [5], physiological mental states, such as fatigue, distraction, vigilance, are obtained more widespread attention, while human reliability, which is usually a concept applying in risk analysis, is less considered in situated driving context. Although physiological mental states can reflect human reliability, for example, high fatigue or low vigilance is related to low reliability of human drivers, they still have some differences as human reliability is considered to be the probability of humans conducting specific tasks with satisfactory performance, while physiological mental states indicates the states of human operators in context.

[1]Chao He, Dirk Söffker are with the chair of Dynamics and Control, University of Duisburg-Essen, Duisburg, Germany chao.he@uni-due.de, soeffker@uni-due.de

A. Physiological mental states

Physiological mental states are widely considered in human operators related safety issues, i.e. vigilance, drowsiness, fatigue, distraction, etc. Five types of measures are commonly used: subjective report measures, biological measures, physical measures, performance measures, and hybrid measures [5]. These measures have their own advantages and disadvantages for physiological mental states measurement. For example, electroencephalography (EEG) of biological measures has highly accuracy rate, but its intrusiveness to drivers limits its application; physical measures in eye movements including PERCLOS, eye closure duration (ECD) are limited by lighting condition (environment), although they are useful non-intrusive measures; some behaviors of human operators are also attractive for physiological mental states estimation, but the reliability and sensitivity of these measures have been questioned as only two or three indicators are applied for the estimation.

B. Human reliability

Human reliability is a more common used concept in probability assessment context, for example, marine engineering [6] and spaceflight application [7]. Human reliability analysis (HRA) is a sophisticated method to calculate human error probability (HEP), it is quantified by the ratio of occurrences of errors to number of opportunities for errors. However, the essential problem in HRA is the possibility of collecting reliability data. As a consequence, many error rates come from expert judgment, which has stunted the application of HRA in industrial safety.

As the likelihood of human error occurrences and the possibilities of gathering relevant data are much more promising in road traffic than other human-in-loop related industry, driving data could be used for HRA. However, reports on reliability of human drivers analysis under situated or dynamic driving context are limited as how to characterize the situated driving context in HRA is less considered. Human reliability will dynamically change in real time with situated driving context so resulting into a situated measure. In this contribution, a modified CREAM approach was established for reliability evaluation of human drivers in situated driving context. The contribution is organized as follows: in Section II , an introduction of CREAM is given. The modified CREAM approach is introduced in Section III, mainly containing how to select new CPCs and calculate human performance reliability score (HPRS). A case study is given in Section IV to illustrate how to evaluate situated human reliability. Finally, conclusions and outlook are provided in Section V.

II. Introduction of CREAM

Human reliability analysis (HRA) is a systematic evaluation method focusing on the analysis, prediction, and prevention of human errors. After years of development, two generations of HRA methods have been established. The so-called 'first generation' of HRA methods is developed based on the idea that human naturally fails to perform tasks because of inherent deficiencies, just like mechanical or electrical components. So human reliability is characterized by the characteristics of the performed tasks [2].

For the so-called 'second generation' of HRA approaches, however, the core assumption is that environment or context is considered as the most important factor affecting human reliability. The widely used methods are a technique human error analysis (ATHEANA) [8], and cognitive reliability and error analysis method (CREAM) [9].

The CREAM approach, developed by Erik Hollnagel in 1998, offers a practical approach for performance analysis as well as attendant prediction. This approach is able to conduct a retrospective analysis of events and a prospective analysis for the design of high-risk systems or process.

A. Contextual control mode

Human cognitive model used in CREAM methodology to model human behaviors is denoted as contextual control mode (COCOM). It is assumed that the degree of control that human operators have on the situation or context is the most important index to estimate human performance and human reliability. Meanwhile, the degree of control can be determined by the context under which human operators perform their tasks. Finally, the degree of control is the core mechanism to determine the relations between context and human reliability.

In CREAM approach, four control modes are established [9].

- Strategic control
  The human operator considers the global context, thus using a wider time horizon. So human operator can have a more efficient and robust performance, which may have a higher reliability.
- Tactical control
  The performance is based on planning, hence more or less follows a known procedure or rule. However, the planning is sometimes limited and too many tasks need to be considered, and may therefore affect reliability more or less.
- Opportunistic control
  The human operator does very little planning or anticipation, perhaps because the context is not clearly understood or because time is too constrained, thus may induce reliability decrease at some extent.
- Scrambled control
  Scrambled control characterizes a situation where there is little or no thinking involved in choosing what to do. In this case, there is a complete loss of situation awareness, and human reliability is very low.

Each control mode corresponds to different human reliability, scrambled control represents the lowest human performance reliability, while strategic control is related to highest human performance reliability. The corresponding HEP interval of each control mode is shown in TableI. The reliability intervals (probability of action failure) of control modes come from statistical data in industries.

TABLE I
CPC control modes and their probability interval

Control modes	**HEP interval**
Strategic mode	(0.00005,0.01)
Tactical mode	(0.001,0.1)
Opportunistic mode	(0.01,0.5)
Scrambled mode	(0.1,1.0)

B. Common performance conditions

Nine common performance conditions (CPCs) are defined as the most significant factors describing the context. These nine CPCs are adequacy of organization, working conditions, adequacy of MMI and operational support, availability of procedures/plans, number of simultaneous goals, available time, time of day (circadian rhythm), adequacy of training and experience, and crew collaboration quality. Each CPC has several different levels, and corresponding expected effect on performance reliability. For example, the CPC of crew collaboration quality has four different levels, which are, very efficient, efficient, inefficient, and deficient, with the corresponding expected effects on performance reliability as improved, not significant, not significant, and reduced, respectively. When the effect on performance reliability of each CPC is identified, CPC score could be determined as ($\sum$reduced, $\sum$improved) where $\sum$reduced represents the sum of reduced effects on performance reliability while $\sum$improved means the sum of improved effects on performance reliability. The effects when CPCs have not significant effects on performance reliability are not considered. The control mode is then identified with a relation map between CPC score and control modes which is shown as Fig.1. For example, if CPC score is (3, 2), it means that 3 reduced and 2 improved effects on performance reliability are identified, respectively. The control mode is then identified as tactical mode.

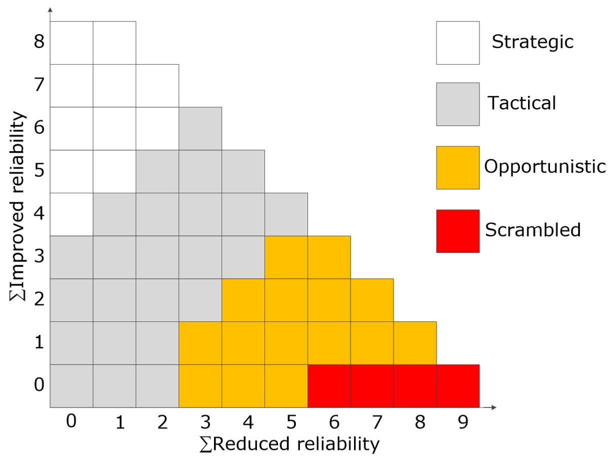

Fig. 1. Relations between CPC score and control modes (adapted from [9])

III. Modified CREAM

The core idea of CREAM considers context as the most crucial factor affecting the human performance failure [2]. Therefore, reliability of human operators could be induced from context in which operators involved. Original CREAM is primarily applied in the human reliability analysis in industry, CPCs in original CREAM can characterize the performance conditions in industry very well. When it is used in other domain, such as driving context, the list of CPCs is not suitable any more, so it is advised to generate a new list of CPCs for the application domain, if the CREAM approach is applied to other domain [10]. In this contribution, the modified CREAM approach is developed in the context of evaluation of human reliability in traffic scenarios.

A. Selection of new CPCs

To characterize the main elements affecting human drivers reliability in situated driving context, a new list of CPCs is proposed. They are number of surrounding vehicles, time to collision (TTC), longitudinal acceleration, lateral acceleration, traffic density, ego-vihicle speed, number of available lanes, actual lane, and general visibility conditions. These CPCs are described by different levels which can be used to assess the expected effect on performance reliability of human drivers. The complete list of CPCs is shown in Table II.

TABLE II
New CPCs and performance reliability

CPC name	Level/description		Expected effect on performance reliability
Number of surrounding vehicles (N)	N=0		Improved
	1≤N≤3		Not significant
	N≤4		Reduced
Time to collision (TTC)	TTC>5.5 s		Improved
	2.5≤TTC≤5.5s		Not significant
	TTC<2.5 s		Reduced
Ego-vehicle speed (V)	V≤22 m/s		Improved
	22<V≤30 m/s		Not significant
	V>30 m/s		Reduced
Longitudinal acceleration (a_{long})	V≤22 m/s	a_{long}≤1.60 m/s^2	Improved
		1.60< a_{long}≤2.32 m/s^2	Not significant
		a_{long} >2.32 m/s^2	Reduced
	22<V≤30 m/s	a_{long}≤1.13 m/s^2	Improved
		1.13< a_{long}≤1.60 m/s^2	Not significant
		a_{long} >1.60 m/s^2	Reduced
	V>30 m/s	a_{long}≤1.13 m/s^2	Not significant
		a_{long} >1.13 m/s^2	Reduced
Lateral acceleration (a_{lat})	V≤22 m/s	a_{lat}≤1.48 m/s^2	Improved
		1.48< a_{lat}≤2.15 m/s^2	Not significant
		a_{lat} >2.15 m/s^2	Reduced
	22<V≤30 m/s	a_{lat}≤1.05 m/s^2	Improved
		1.05< a_{lat}≤1.48 m/s^2	Not significant
		a_{lat} >1.48 m/s^2	Reduced
	V>30 m/s	a_{lat}≤1.05 m/s^2	Not significant
		a_{lat} >1.05 m/s^2	Reduced
Traffic density (D)	Low (D≤7)		Improved
	Medium (8≤D≤14)		Not significant
	High (D≥15)		Reduced
Number of available lanes	Three lanes		Improved
	Two lanes		Not significant
	One lane		Reduced
Actual lane	The first lane (The right lane)		Improved
	The second lane (The middle lane)		Not significant
	The third lane (The left lane)		Not significant
General visibility conditions	Daytime with sunny weather		Improved
	Morning or nightfall with sunny weather		Not significant
	Evening or foggy or rainy or snowy		Reduced

- Number of surrounding vehicles
  The behavior of ego-vehicle is affected by surrounding vehicles as driving context could be more complex when more vehicles are surrounded. Based on literature [11], surrounding vehicles can be defined as vehicles that the time to collision (TTC) of front/rear vehicle, and vehicles in the adjacent lanes to ego-vehicle is less than 1.5 s. If ego-vehicle is not surrounded by any vehicles, human drivers are not distracted by surrounding vehicles. In this case, the expected effect on performance reliability is improved. When ego-vehicle is surrounded by 1-3 vehicles, the abilities of human drivers are just right for this situation. When more than 3 surrounding vehicles exist, it seems to have a reduced effect on performance reliability.
- Time to collision (TTC)
  Time to collision (TTC) is an important parameter indicating the time it would take a following vehicle to collide with a leading vehicle [12]. This parameter can be used to characterize the safety of vehicle following and lane changing. When TTC $\geq$ 5.5 s, human drivers have enough time to complete different operations, like lane changing or braking, so the effect on performance reliability is improved. Evidence from [13] has presented that TTC of 2.5 s could be regarded as a minimum value that should be avoided in normal traffic conditions. When TTC $\leq$ 2.5 s, abilities of drivers to handle the situation are insufficient, so a reduced effect is generated.
- Ego-vehicle speed

Ego-vehicle speed, as an important index to characterize driving behavior, is closely related to driving safety. Some physiological properties of human drivers, like visual ability and reaction time, are easily affected by vehicle speed, and the performance reliability of human drivers is then influenced by physiological properties. In this contribution, three levels of speed are identified, speed larger than 110 km/h, speed between 80 km/h and 110 km/h, and speed less than 80 km/h, and their corresponding effects on performance reliability are reduced, not significant, and improved.

- Longitudinal acceleration
  Acceleration is fundamental to define the behavior of drivers as it describes the motion of vehicles. Acceleration, which can be used to classify drivers' behaviors as safe or unsafe [14], can be divided into longitudinal and lateral acceleration. Acceleration is closely related to driving speed for safety driving issues, acceleration should decrease when vehicle is in high speed. The relationship between longitudinal acceleration and vehicle speed is concluded in [14], [15]. So the longitudinal acceleration corresponding to different driving speed is also obtained.
- Lateral acceleration
  The relationship between longitudinal acceleration and lateral acceleration is explained in [14], as the longitudinal acceleration is 0.925 times the lateral acceleration.
- Traffic density
  Traffic density expresses the average number of vehicles that occupy one kilometer of traffic lane. Driving behavior of human driver is affected by traffic density. When traffic density is low, traffic context is relatively simple, drivers have more operating options for situations encountered, therefore, relatively high performance reliability of human drivers is reached. On the contrary, available options for human drivers are limited and uncertainty situations will also increase when traffic density is high. Meanwhile, higher traffic (approximately 15 vehicles per kilometer) could result in higher workload and demand compared to low traffic density situations (approximately 7 vehicles per kilometer) [16], [17]. Considering identified TTC and ego-vehicle speed in this contribution, traffic density can be classified into three levels, namely, low traffic density (less than 7 vehicles per kilometer), medium traffic density (between 8 to 14 vehicles per kilometer), and high traffic density (more than 15 vehicles per kilometer), which corresponds to the effects to performance reliability as improved, not significant, and reduced.
- Number of available lanes
  Number of available lanes illustrates the complexity of traffic context. More available lanes give drivers more safety redundancy that drivers have more choices when emergency events are encountered.
- Actual lane
  Following the traffic rules in Germany, vehicles should keep driving at the right lane. It is only allowed to overtake from left lane. Vehicles need to return to the right lane after overtaking, long-term use of the left lane is not allowed. These traffic rules will affect driving behaviors of drivers in different lanes.
- General visibility conditions
  General visibility conditions affect perception level of human drivers on surrounding context. With low level of conditions, many context information could not be perceived by human drivers, which may have high risk on vehicle driving. General visibility conditions are mainly influenced by the time of the day and weather conditions.

B. Definition and calculation of new human performance reliability score

In original CREAM, after the identification of levels of CPCs, CPC score could be determined as ($\sum$reduced, $\sum$improved). The control mode and related HEP could be, therefore, identified. This method is valid for the assessment of operation as a whole, or major segments of the operation, when is for human operation in situated context, it becomes invalid. Human performance reliability is constantly changing with time as situated context is encountered, a new evaluation system for the reliability of human operators considering the time of operation, therefore, needs to be proposed.

In this contribution, a new concept of human performance reliability score (HPRS) is introduced to define the continuously calculated performance reliability of human operators in dynamic context. The equation is

$$HPRS = \lambda_1 \cdot \sum reduced + \lambda_2 \cdot \sum improved, \quad (1)$$

where λ denoting related weights. Here λ = 1, denotes improving effects, which λ = -1 reducing effects.

The CPC score could be used to build the relations between HPRS and control modes. From Fig.1, control modes of human operations could be identified by CPC score of ($\sum$reduced, $\sum$improved) as the measurement method. At the same time, HPRS is also closely related to CPC score. Therefore, when CPC score is determined, relations between HPRS and control modes are also obtained. Some examples can be provided to illustrate how to build the relations. If CPC score is identified as (1,6) which is related to control mode of strategic, HPRS can be obtained as 5 which is also determined as strategic control; when CPC score changes to (2,3), it means that control mode can be determined as tactical, and HPRS is then obtained as 1 which can be also determined as tactical control. According to this process, each CPC score in Fig.1 can be converted into HPRS which has the same control mode with the corresponding CPC score. CPC scores of (7,1) and (8,1) are excepted as their corresponding HPRS are classified as scrambled level, although their CPC scores are in opportunistic mode.

The HPRS can be therefore identified into four levels based on control mode. They are strategic level (4≤HPRS≤9), tactical level (-1≤HPRS≤3), opportunistic level (-5≤HPRS≤-2), and scrambled level (HPRS≤-6), in which strategic level has the highest reliability, and scrambled level has the lowest reliability. In the same levels, larger HRS means higher reliability. In this case, the performance reliability of human operators in every time spot could be identified, so performance reliability of human operators could be evaluated continuously with time. The relations between HPRS and control modes over time is shown as Fig.2.

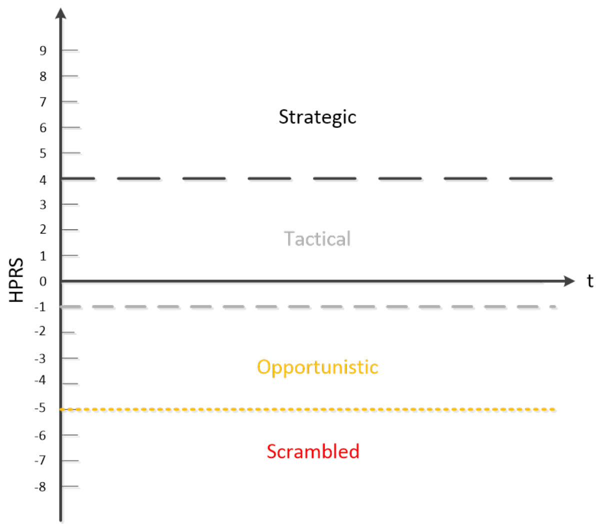

Fig. 2. Relations between HPRS and control modes

IV. Illustrating Experimental Example

A. Experimental setup and procedures

A professional driving simulator SCANeR^{TM} studio as shown in Fig.3 is applied to collect driving data. The simulator consists of five monitors, base-fixed driver seat, steering wheel, and pedals. The three rear mirrors displaying on the corresponding positions of the monitors could help participants understand the situations behind the ego-vehicle.

Fig. 3. Driving simulator, Chair of Dynamics and Control, U DuE

Driving scenarios are set on a three-lane dual carriage highway. Highway has its own characteristics differing from other roads. Highway is usually in less changed road conditions. This feature could induce some driving issues, for example, the braking distance will be extended with high speed driving. It is also easy to be fatigue with the monotonous road conditions. So the levels for the assessment of highway features are different from other driving scenarios. For instance, vehicle speed with 110 km/h is allowed in highway, but it is not allowed in urban roads.

B. Case analysis

In driving process, HPRS may change with time when different driving operations have been performed to cope with the situated context. HPRS may be at strategic level for a long time, or occasionally at tactical level if drivers' competence is sufficient for the situations. On the contrary, HPRS may be lastingly at opportunistic level, even scrambled level if drivers are lack of experiences and cannot cope with driving situations. To fully describe the four different levels of HPRS in driving process, an artificial case is introduced as Fig.4.

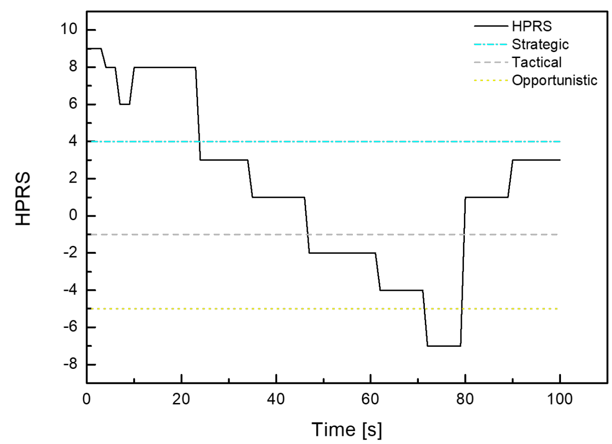

Fig. 4. Case study of artificial HPRS

From Fig.4, it becomes obvious that HPRS changes in four different control levels over time. When HPRS is in strategic and tactical levels, which means the performance of human drivers is efficient and robust, so human drivers have high reliability on the situations. It can be considered that human drivers are lack of understanding of the situations because of negative physiological mental states, or the time is constrained when HPRS is in opportunistic mode. In this case, some actions should be taken to get human drivers back to the loop, for instance, steering wheel vibration, or audio warning. Takeover operation should be taken by assisted system as which has higher reliability than human drivers when HPRS is in scrambled level.

C. Experimental results analysis

The actual data collected by driving simulator are processed and the expected effect on performance reliability of each new generated CPC is evaluated. The experimental result of HPRS with a participant in a specific highway scenario is shown as Fig.5. From Fig.5, it can be obtained that in the whole process of driving, except for a very short period of time, HPRS is in tactical level, during most of of the driving time, HPRS is in strategic mode, which indicates high reliability of the driver coping with the situations.

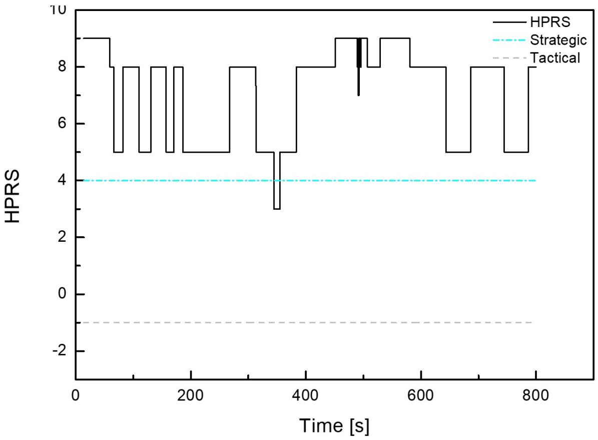

Fig. 5. Experimental result of actual HPRS

V. Conclusions and Outlook

In this contribution, a new approach defining a situated and dynamical human reliability measure is established. The approach is based on the well-known CREAM approach, which is modified with respect to the use in dynamical context. Based on the introduction of control modes and CPCs in original CREAM, the core idea of CREAM is concluded as environment or context is the most important factor affecting human reliability. So, it is reasonable to generate a new list of CPCs to characterize other application domains. Therefore, a new list of CPCs for situated driving context is generated, their corresponding levels and effects on performance reliability are also identified. A concept of HPRS, which can be evaluated by control modes, is proposed for continuously evaluation of human performance reliability over time. A case study is introduced to explain how this approach is applied for the evaluation of performance reliability of human drivers, and the application of actual driving data indicates the applicability and correctness of the proposed method.

In the next step, more participants will be involved in the experiments and different scenarios will be tested for reliability evaluation of human drivers.

Acknowledgment

The research reported in this paper is partly supported by China Scholarship Council through scholarship received by the first author for his Ph.D. study at the Chair of Dynamics and Control, UDE, Germany.

References

[1] T. Trager Jr, "Case study report on loss of safety system function events," *Report No. AEOD C*, vol. 504, 1985.

[2] Q. Zhou, Y. D. Wong, H. Xu, V. Van Thai, H. S. Loh, and K. F. Yuen, "An enhanced cream with stakeholder-graded protocols for tanker shipping safety application," *Safety science*, vol. 95, pp. 140–147, 2017.

[3] S. Shappell, C. Detwiler, K. Holcomb, C. Hackworth, A. Boquet, and D. A. Wiegmann, "Human error and commercial aviation accidents: an analysis using the human factors analysis and classification system," in *Human Error in Aviation*, pp. 73–88, Routledge, 2017.

[4] S. Singh, "Critical reasons for crashes investigated in the national motor vehicle crash causation survey," tech. rep., 2015.

[5] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," *IEEE transactions on intelligent transportation systems*, vol. 12, no. 2, pp. 596–614, 2010.

[6] S.-T. Ung, "A weighted cream model for maritime human reliability analysis," *Safety science*, vol. 72, pp. 144–152, 2015.

[7] J. Calhoun, C. Savoie, M. Randolph-Gips, and I. Bozkurt, "Human reliability analysis in spaceflight applications," *Quality and Reliability Engineering International*, vol. 29, no. 6, pp. 869–882, 2013.

[8] S. Cooper, A. Ramey-Smith, J. Wreathall, and G. Parry, "A technique for human error analysis (atheana)," tech. rep., Nuclear Regulatory Commission, 1996.

[9] E. Hollnagel, *Cognitive reliability and error analysis method (CREAM)*. Elsevier, 1998.

[10] H. Taga, K. Furuta, and T. Kanno, "Human reliability analysis of car drivers in urban intersections," *Cognition, Technology & Work*, vol. 14, no. 4, pp. 365–377, 2012.

[11] F. Bella and R. Russo, "A collision warning system for rear-end collision: a driving simulator study," *Procedia-social and behavioral sciences*, vol. 20, pp. 676–686, 2011.

[12] V. E. Balas and M. M. Balas, "Constant time to collision platoons," *Int. J. of Computers, Communications and Control*, vol. 3, pp. 33–39, 2008.

[13] R. Van Der Horst and J. Hogema, "Time-to-collision and collision avoidance systems," *6th ICTCT workshop Salzburg, Proceedings*, pp. 129–143, 1993.

[14] L. Eboli, G. Mazzulla, and G. Pungillo, "Combining speed and acceleration to define car users' safe or unsafe driving behaviour," *Transportation research part C: emerging technologies*, vol. 68, pp. 113–125, 2016.

[15] A. S. Zeeman and M. J. Booysen, "Combining speed and acceleration to detect reckless driving in the informal public transport industry," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, pp. 756–761, IEEE, 2013.

[16] C. Gold, M. Körber, D. Lechner, and K. Bengler, "Taking over control from highly automated vehicles in complex traffic situations: the role of traffic density," *Human factors*, vol. 58, no. 4, pp. 642–652, 2016.

[17] N. A. Stanton and M. S. Young, "Driver behaviour with adaptive cruise control," *Ergonomics*, vol. 48, no. 10, pp. 1294–1313, 2005.