Data driven-based human reliability analysis for individualized human supervision and reliability evaluation in situated context

Von der Fakultät für Ingenieurwissenschaften,
Abteilung Maschinenbau und Verfahrenstechnik
der
Universität Duisburg-Essen
zur Erlangung des akademischen Grades
eines
Doktors der Ingenieurwissenschaften
Dr.-Ing.
genehmigte Dissertation

von

Chao He
aus
Hubei, China

Gutachter: Univ.-Prof. Dr.-Ing. Dirk Söffker
Prof. Dr. Hyungju Kim

Tag der mündlichen Prüfung: 04. Mai 2023

To my parents and sisters

Acknowledgements

This paper is the result of my research work, which I conducted at the Chair of Dynamics and Control (SRS) at the University of Duisburg-Essen from 2018 to 2022. I would like to thank all the people who have supported and guided me during this time.

First and foremost, I would like to express my sincere gratitude to my supervisor, Univ.-Prof. Dr.-Ing. Dirk Söffker, who offers me the opportunity to be his student. I thank him for his continuous guidance, valuable support, and patience throughout my Ph.D. study.

I am also grateful to my second supervisor, Associate Professor Dr. Hyungju Kim, for his valuable feedback, scientific questions, and helpful suggestions to improve this work.

I would like to thank all my colleagues at the Chair of Dynamics and Control (SRS) for their kind support, encouragement, and cooperation.

Specially, I would like to thank my father, Defu He, my mother, Mingzhen Hou, who raise me up and teach me to be kind and brave. I thank my old sisters, Yan He and Jing He, who always encourage and believe in me.

Mülheim an der Ruhr, May 2023 Chao He

Kurzfassung

Durch die fortschreitende Automatisierung in der Industrie sinkt die Zahl der Unfälle, jedoch steigt die Anteil der Unfälle, die auf den menschlichen Faktor zurückzuführen sind am Gesamtunfallgeschehen. Die Zuverlässigkeit des Menschen wird zur Schlüsselfrage in Mensch-Maschine-Systemen, insbesondere bei sicherheitsrelevanten Aufgaben und Operationen.

Es wurden verschiedene Methoden der menschlichen Zuverlässigkeitsanalyse (Human Reliability Analysis, HRA) entwickelt, die systematisch zur Analyse, Vorhersage und Vorbeugung von menschlichen Fehlern eingesetzt werden können. Es lassen sich drei grundlegende Lücken in den bestehenden HRA-Methoden feststellen: i) das Fehlen von Möglichkeiten zur Erhebung relevanter Daten in der Human-in-Loop-Branche, ii) die fehlende Berücksichtigung der dynamischen menschlichen Zuverlässigkeit im situierten Kontext und iii) die starke Abhängigkeit von Expertenwissen im Bewertungsprozess. Die Zuverlässigkeit des menschlichen Fahrers im dynamischen Kontext steht im Mittelpunkt dieser Arbeit, daher werden ohne Einfluss auf die Verallgemeinerbarkeit menschliche Fahrer in sich dynamisch verändernden Situationen aufgrund der leichten Verständlichkeit als Beispiel verwendet. Mit der Entwicklung der Überwachung des menschlichen Fahrerverhaltens im dynamischen Fahrkontext werden Fahrverhaltensdaten generiert, die für die HRA genutzt werden können.

In dieser Arbeit wird der modifizierte CREAM-Ansatz (Cognitive Reliability and Error Analysis Method) zur Bewertung der menschlichen Fahrleistung angewendet. Um die Abhängigkeit von Expertenwissen zu verringern, werden die Leistungsstufen im modifizierten CREAM-Ansatz durch automatisches Clustering der Daten bestimmt. Drei Daten-Clustering-Ansätze, darunter FN-DBSCAN (Fuzzy neighborhood density-based spatial clustering of application with noise), CLUSTERDB* und GMFPE (genetic-based membership function parameter estimation), werden auf Fahrdaten angewandt, um die Parameter der Zugehörigkeitsfunktionen zu definieren, die den Leistungsstufen im modifizierten CREAM-Ansatz entsprechen. Als Ergebnis wird ein neuer Ansatz entwickelt, der die dynamischen Aspekte der menschlichen Zuverlässigkeit berücksichtigt. Das Konzept und der Entwurf einer neuen Bewertung als Human Performance Reliability Score (HPRS) als Funktion der Zeit wird für die quantitative und dynamische Bewertung der individuellen menschlichen Leistung vorgeschlagen. Die HPRS-Ergebnisse mit verschiedenen Clustering-Ansätzen werden für denselben Zeitraum verglichen.

Um die kritischen Verhaltensweisen beim situierten Fahren zu erkennen und zu bewerten, muss eine Quantifizierung der menschlichen Verhaltensebenen vorgeschlagen werden. Dies ist vor allem daher wichtig, da die Fahrzeugautomatisierung im Fall der Übernahme der Fahrzeugführung sowohl auf untrainierte Personen trifft wie

auch die Personen dann unter Zeitdruck handeln müssen. Das SRK-Modell (Skill-Rule-Knowledge) von Rasmussen ist im Bereich der menschlichen Faktoren sehr bekannt. Ebenso ist bekannt, dass auf Fähigkeiten basierende Verhaltensweisen die höchste menschliche Zuverlässigkeit aufweisen, während wissensbasierte Verhaltensweisen mit den niedrigsten Zuverlässigkeitswerten verbunden sind. Zwar gibt es zahlreiche Studien zur menschlichen Fehlerwahrscheinlichkeit (HEP), die in der Regel direkt oder indirekt auf diese drei Verhaltensebenen zurückgeführt werden, aber eine kohärente, konsistente Darstellung, insbesondere unter Verwendung von Datenquellen, ist bisher nicht verfügbar. In dieser Arbeit wird die Quantifizierung menschlicher Verhaltensebenen mit dem SRK-Modell von Rasmussen anhand von drei Datenbanken vorgenommen. Die Auswirkungen von Zeitdruck und Training auf die menschliche Zuverlässigkeitsvermittlung werden ebenfalls auf der Grundlage einschlägiger Veröffentlichungen analysiert.

Zur Bestimmung des HEP dieser drei Ebenen werden drei Datenbanken, die Technik zur Vorhersage der menschlichen Fehlerrate (THERP), die Savannah River Site Human Reliability Analysis (SRS-HRA) und die Nuclear Action Reliability Assessment (NARA), aus den Methoden der Human Reliability Analysis (HRA) verwendet. Das Verfahren umfasst die Identifizierung der Aufgaben einschließlich des beteiligten Bedieners und der von den Analysten getroffenen Annahmen sowie die Einordnung der Aufgaben in den geeigneten kognitiven Verhaltensmodus (CBM). In diesem Fall wird die Beziehung zwischen SRK-Niveau und menschlicher Fehlerwahrscheinlichkeit (HEP) abgebildet. Die Auswirkungen der beiden im Automatisierungskontext sehr relevanten Performance Shaping Factors (PSFs), Zeitdruck und Trainings-/Wissensabbau, auf das menschliche Verhalten beim Schalten werden analysiert und die Erklärungen für das SRK-Schalten vorgestellt. In diesem Fall wird eine allgemeinere Struktur erstellt, um das dynamische Verhalten von Ebenen zu veranschaulichen, die unter verschiedenen Bedingungen in sechs Richtungen schalten. Aus den Ergebnissen schließen wir, dass die Ebenen der Fähigkeiten, Regeln und des Wissensverhaltens in Bezug auf HEP kontinuierlich sind und daher einen neuen Einblick in diesen Schlüsselaspekt der Quantifizierung menschlicher Faktoren ermöglichen. Basierend auf dieser Analyse werden die Folgen der Alltagsautomatisierung im Kontext autonomer Verkehrssysteme in Kombination mit menschlicher Qualifikation und Zuverlässigkeitsverschlechterung aus dieser spezifischen und in der aktuellen Automatisierungsdiskussion sehr intensiv diskutiert.

Der vorgeschlagene Ansatz wird es ermöglichen, künftige Automatisierungssysteme mit Warnung, Unterstützung oder Umschaltung auf vollautomatische Steuerung zu etablieren, um menschliche Fehler zu vermeiden. Die vorgestellte Diskussion über die Verknüpfung von SRK-Ebenen und HEP bietet eine neue Perspektive auf die absehbaren Folgen einer weiteren Automatisierung in Anwendungsbereichen mit zunehmender Automatisierung alltäglicher Aufgaben (z. B. bei der Nutzung eines hochautomatisierten Fahrzeugs).

Abstract

Human factor-related accidents account for an increasing portion of the total accidents through the advancing level of system automation. Human reliability becomes the key issue in human-machine systems especially for safety-relevant tasks and operations.

Various human reliability analysis (HRA) methods to systematically incorporate for the analysis, prediction, and prevention of human errors have been developed. Three fundamental gaps in the existing HRA methods can be stated: i) the lack of possibilities of gathering relevant site data in human-in-loop related industry, ii) the missing consideration of dynamic human reliability in situated context, and iii) the deep reliance on expert knowledge in evaluation process. Human operator's reliability in dynamic context is the focus of this thesis, therefore without loss of generality human drivers in dynamically changing situations are used as example case due to easy comprehensibility. With the development of human driver behavior monitoring in dynamic driving context, driving behavior data are generated and could be used for HRA.

In this thesis, the modified CREAM (cognitive reliability and error analysis method) approach is applied for the evaluation of human driver performance. To reduce the reliance on expert knowledge, the performance levels in the modified CREAM approach are determined by automated data clustering. Three data clustering approaches including FN-DBSCAN (fuzzy neighborhood density-based spatial clustering of application with noise), CLUSTERDB*, and GMFPE (genetic-based membership function parameter estimation) are applied to driving data defining the membership function parameters which are corresponding to the performance levels in the modified CREAM approach. As result a new approach addressing dynamically aspects for human reliability is developed. The concept and the design of a new evaluation of human performance reliability score (HPRS) as a function of time is proposed for the quantitative and dynamic evaluation of individualized human performance. The HPRS results with different clustering approaches for the same time period are compared.

To detect and evaluate the critical behaviors in situated driving, the quantification of human behavior levels needs to be proposed. Rasmussen's SRK (skill-rule-knowledge) model is well known in the field of human factors. Likewise, it is well known that skill-based behaviors have the highest human reliability, while knowledge-based behaviors are associated with the lowest reliability scores. Although numerous studies exist on human error probability (HEP), correspondingly typically attributed directly or indirectly to these three levels of behavior, a coherent, consistent representation, especially using data sources, has not been available. In this thesis, the quantification of human behavior levels with Rasmussen's SRK

model is given based on three databases. Effects of time pressure and training on human reliability switching are also analyzed based on related publications.

To determine the HEP of these three levels, three databases, technique for human error rate prediction (THERP), Savannah river site human reliability analysis (SRS-HRA) and nuclear action reliability assessment (NARA), from human reliability analysis (HRA) methods are used. The procedure contains identifying the tasks including the operator involved and the assumptions the analysts made and classifying the tasks into suitable cognitive behavior mode (CBM). In this case, the relationship between SRK levels and HEP is mapped. The effects of the two in automation context very relevant performance shaping factors (PSFs), time pressure and training/knowledge degradation, on human behavior levels switching are analyzed and the explanations of the SRK switching are presented. In this case, a more general structure is established to illustrate the dynamic behavior of levels switching with six directions under different conditions. From the results we conclude that skill, rule, and knowledge behavior levels are continuous in terms of HEP and therefore allow a new inside into this key aspect of human factor quantification. Based on this analysis the consequences of daily automation in the context of autonomous transport systems in combination with human qualification and reliability degrading is from this specific and in the current automation discussion very intensively discussed.

The proposed approach will allow future automation systems including warning, assistance, or situated switch over to fully automated control to be established for the avoidance of human errors. Meanwhile, the presented discussion linking SRK levels and HEP gives a new perspective on the foreseeable consequences of further automation in application areas with increasing automation of everyday tasks (like using a highly automated vehicle).

Contents

List of Figures

List of Tables

Nomenclature

Symbols

t	Time
$HEP_{dynamic}$	Dynamic HEP
HEP_{static}	Static HEP
$HEP_{nominal}$	Nominal or basic HEP
PSF	Performance shaping factors
$BHEP$	Basic HEP
HFE	Human failure events
EFC	Error forcing context
S	Scenario
UA	Unsafe action
A	Fuzzy set
$\mu_A(x)$	Degree of membership function of x in A
$N_x(y)$	Fuzzy neighborhood membership function
$d(x, y)$	Distance between any points x and y
ϵ	Maximal threshold of the distance between points
k	Parameter to depict neighborhood membership functions
$FN(x; \epsilon_1)$	Fuzzy neighborhood set of point $x \in X$ with parameter ϵ_1
ϵ_1	Minimal threshold of the neighborhood membership degree
$cardFN(x; \epsilon_1, \epsilon_2)$	Fuzzy core point
ϵ	Average distance between adjacent data
d^{max}	Maximum distance between any points
$DB*(nc)$	Overall similarity of all cluster nc
d_{ip}	Distance between the i-th and p-th centroid
S_i	Scatter distance
C_i	The i-th cluster
$diff$	Distance between adjacent data points
s_i	Similarity value
N	Parameter to decide the membership functions shape
σ_s	Standard deviation of all $diff$
$size$	Total size of chromosomes of all variables
m_i	Number of membership functions of variable i
x_l, x_r	Trapezoidal core parameters
x_a, x_e	Trapezoidal support parameters
$HPRS$	Human performance reliability score
$reduced$	Reduced effects on performance reliability
$improved$	Improved effects on performance reliability
λ	Weighting values

Abbreviations

ADAS	Advanced driver assistance system
KSS	Karolinska sleepiness scale
NASA-TLX	NASA task load index
MCH	Modified Cooper Harper scales
SWAT	Subjective workload assessment technique
VACP	Visual auditory cognitive and psychomotor
MRQ	Multiple resources questionnaire
EEG	Electroencephalography
ECG	Electrocardiogram
EOG	Electro-oculography
sEMG	Surface electromyogram
HRA	Human reliability analysis
HEP	Human error probability
SRK	Skill-, rule-, knowledge-
CREAM	Cognitive reliability and error analysis method
CPCs	Common performance conditions
HPRS	Human performance reliability score
GEMS	Generic error-modeling system
PSFs	Performance shaping factors
HEART	Human error assessment and reduction method
EPCs	Error producing conditions
THERP	Technique for human error rate prediction
SLIM-MAUD	Success likelihood index method using multi-attribute utility decomposition
SPAR-H	Standardized plant analysis risk HRA
ATHEANA	A technique for human event analysis
HuRECA	Human reliability evaluator for control room actions
ASEP	Accident sequence evaluation program
HCR	Human cognition reliability
HFEs	Human failure events
UAs	Unsafe actions
EFC	Error forcing context
NARA	Nuclear action reliability assessment
GTTs	Generic task types
APOA	Assessed proportion of affect
IDAC	Information decision and action crew
SA	Situation awareness
ADS-IDAC	Accident dynamics simulator with the information decision and action in a crew context operator model
COSIMO	Cognitive simulation model
CREATE	Cognitive reliability assessment technique

DYLAM	Dynamic logic analysis method
GEMS	Generic error modeling system
HERMES	Human error reliability methods for event sequences
OAT	Operator action tree
SRS-HRA	Savannah river site HRA
NPP	Nuclear power plant
EF	Error factor
COREDATA	Computerized operator reliability and error data
COCOM	Contextual control mode
FN-DBSCAN	Fuzzy neighborhood density-based spatial clustering of application with noise
GMFPE	Genetic-based membership function parameter estimation
TTC	Time to collision
SOM	Situation-operator-modeling

1 Introduction

With the development of automation, not only simple and repetitive actions are replaced by machines, more and more decision-making tasks have also begun to rely on the assistance of machines. Most safety-critical systems or fields such as power plants in energy production [ZTC+17], guiding or flying aircrafts in aviation [DBD17], or in transportation in general [WS18], automation is involved. Automation has profoundly influenced human behaviors in human-machine-systems, as many repetitive, many mechanical tasks can now be left entirely to machines, humans mainly are related to supervisory control tasks. While in safety critical fields typical automation side-effects firstly very intensively discussed by [Bai83] are considered and well-known this is not necessarily the case in other fields. At the same time, higher levels of automation are increasingly capable of performing tasks that were previously thought to be performed only by humans. Automation has been varied to different levels, with higher levels representing increased machine autonomy. In [SV78], the automation of decision and action selection is divided into ten levels where level 1 indicates no assistance from the automation and level 10 presents that the automation decides everything and human is ignored. Humans cannot be excluded except for the automation of level 10. The society of automotive engineering (SAE) defines six levels of automation regarding driving from level 0 of no automation to level 5 of full automation [Shu19]. It demonstrates that the human driver is not able to be decoupled with driving activities even with full automated vehicle as the driver still needs to monitor the driving situations and possibly to takeover the vehicle. In high-risk environment such as air traffic conflict prediction, decision automation should be set that allowing operator input into the decision-making process. Meanwhile, the additional time required for humans to decide how to respond to an automated situation assessment may impose unsafe events. For example, different takeover time are critical to the reliability of automated vehicle drivers in dealing with emergencies [WS18]. Automation can have both beneficial and negative effects on human performance [PSW00]. When changes in environmental or system states are controlled by another agent, humans tend to be less aware of them indicating the operator's situation awareness of the dynamic features of the working environment is reduced. If the system functions are consistently performed by automation, humans will not be as skilled in performing the functions which means skill degradation.

The role played by humans is gradually shifting from active control to passive monitoring in human-machine systems [MSE+21]. Human supervisory control could be explained as interaction with a computer/automation system to transform data or to produce control actions [ST92]. From [She21], the roles played by human supervisor include: i) planning offline what the task to do and how to do; ii) programming the computer/automation what has to be executed; iii) monitoring the automatic actions online to make sure all actions are going as planned and occurring failures are detected; iv) intervening to take over the control of the automation when the desired goal has been reached or emergency situations happen; v) learning from experience to perform better in the future. Therefore, humans are still vital in operation process in human-automation system. Human-related accidents accounts for the highest proportion of total accidents in various fields. The proportion of human error-related accidents in several industries and activities are collected in Table 1.1.

Table 1.1: Proportion of human error-related accidents

Accident	Proportion (%)	Year	Reference
Nuclear power facilities	50 - 70	1985	[TJ85]
Maritime	80	2017	[ZWX+17]
Aviation	60 - 80	1996	[SDH+17]
Chemical industry	63	2006	[PP06]
Light vehicles	94	2015	[Sin15]
Heavy truck	80	2007	[Dhi07]

1.1 Motivation and objectives of the work

Human performance is the key to systems safety. Many technologies and measurements are developed to monitor and assist human behaviors in different application fields, such as the advanced driver assistance systems (ADAS) in driving context and physiological metal states monitoring and evaluation including vigilance, fatigue, distraction, etc. In [DHUM10], the mainly existing measurement methods for human driver inattention are summarized and categorized into five groups, which are subjective report, biological measures, physical measures, performance measures, and hybrid measures. In Table 1.2 the applied methods and their corresponding advantages and disadvantages are presented. In aviation, the predictions of situation awareness of pilots are improved with human performance model [HGW+11]. In maritime, the impact of seafarers' emotion on their performance is investigated with EEG and self rating [FZBD+18].

In situated and dynamic context, the timing of the operator taking over the task as the automation system is failed to complete the task or the automation intervening into the operation process because of critical human performance is important. In this case, a method to evaluate human performance in real time is needed and the criteria to define the critical human performance is necessary to trigger the automation or assisted system when emergencies occur. The quality of human performance also needs to be taken into account when it comes to taking over an automated system that has failed. Human performance, therefore, needs to be evaluated quantitatively in real time in situated context.

Table 1.2: Existing measurements in driving scenario (adapted from [DHUM10])

Measurement	Methods	Advantages	Disadvantages
Subjective report	KSS; NASA-TLX; MCH; SWAT; VACP; MRQ	Easy to quantify the results; Easy to process	Hard to design; Strong subjectivity; Not possible in real time
Biological measures	EEG; ECG; EOG; sEMG	Highly accuracy rate; Reliable	Highly intrusive; Not convenient
Physical measures	Eye movement (PERCLOSS; eye closure duration; blink duration; eye closure speed); Mouth activities (Lip features: normal; yawning; talking); Head pose	Non-intrusive; Easy to use	Limited by environment; Lighting condition
Performance measures	Steering wheel movement (angle); Forces on the pedals; Vehicle velocity; Accelerator pedal position; Vehicle trajectory	Non-intrusive	Accuracy rate varies between individuals
Hybrid measures	Eye gaze, head orientation, heart rate; head orientation and surrounding salience map; Eye gaze, blink, and environment parameters	Highly accuracy rate; Reliable	Complicated

KSS: Karolinska sleepiness scale; NASA-TLX: NASA task load index; MCH: Modified Cooper Harper scales; SWAT: Subjective workload assessment technique; VACP: Visual auditory cognitive and psychomotor; MRQ: Multiple resources questionnaire. EEG: Electroencephalography; ECG: Electrocardiogram; EOG: Electro-oculography; sEMG: surface electromyogram.

Human reliability analysis (HRA) provides a well-structured framework to evaluate human performance qualitatively and quantitatively. Human error probability (HEP) calculated with HRA methods could be the reference to determine whether the takeover action is required. Human reliability analysis methods have been propsoed to systematically incorporate for the analysis, prediction, and prevention of human errors. However, the existing HRA methods can not be realized in real time in situated context as most of the existing HRA methods are established for static tasks analysis as the progressions of event are not considered. In this case, a new approach to evaluate human reliability quantitatively in real time in situated context needs to be established.

In this thesis, a new and dynamic human reliability evaluation approach for situated context is established based on cognitive reliability and error analysis method (CREAM) [Hol98], fuzzy theory, and three different data clustering approaches. This new approach could evaluate individual human performance reliability in real time. Most of the existing HRA methods could be considered as static HRA methods as events are analyzed for an assumed window of time and the event evolution is not considered. Moreover, these methods are mainly applied to crews in industrial factories and nuclear power plants, when considering human performance reliability of the individual, the existing HRA methods cannot be adopted properly. With the new concept of human performance reliability score (HPRS) proposed in this approach, human performance reliability could be evaluated with time on second timescales, indicating that human performance reliability is evaluated dynamically. Meanwhile, individual reliability is evaluated with HPRS. Human performance data are clustered with different data clustering approaches, the grouped data represent operators' operational characteristics, or more directly, the human experience regarding different situations. Operators prefer to respond to situations with familiar operation behaviors/ stored rules/ skilled actions, while those rarely occurring behaviors indicate that operators' experience with situations is insufficient, denoting that the human performance reliability could be deduced from the grouped data.

When human performance reliability is calculated, how to detect and evaluate the critical behavior needs to be determined. In this case, the levels of skill-, rule-, and knowledge-based behavior (SRK) [Ras83] framework are quantified. With the classification of levels in three HRA databases, the HEP intervals of SRK levels are determined. In original CREAM approach, four control modes with different HEP intervals are provided, but these HEP interval values are defined by expert knowledge, the connection between CREAM approach and the SRK framework needs to be established with the comparison of HEP values in these two methods, therefore, the structure of HPRS with SRK levels could be defined. The HPRS results could finally be evaluated and the critical behaviors could be detected with SRK levels.

The contributions of this thesis could be summarized as following points.

- A new human reliability evaluation method for situated context is established, the individual human performance reliability is quantified with the new concept of HPRS.
- The HEP intervals of three levels (skill-, rule-, and knowledge-based) in SRK framework is quantified with HRA databases.
- The levels to evaluate HPRS results are defined with the comparison of control modes in CREAM approach and the three levels in SRK framework. In this case, the critical behaviors could be detected.

In addition, some research gaps of this thesis are discussed in chapter 2.3.

1.2 Outline of the thesis

Human reliability needs to be monitored and evaluated in situated context. The goal of this work is to realize the monitoring and evaluation of human reliability quantitatively in situated context online.

In chapter 2, the background knowledge of HRA methods is reviewed, including the basic concept involved in HRA methods and the development of HRA methods. The research gaps of the existing HRA methods is summarized. In chapter 3, with the introduction of cognition process in HRA methods and the well-known skill-rule-knowledge (SRK) framework, the quantification of SRK levels is presented and the effects of time pressure and training on the levels switching are discussed. In chapter 4, a modified fuzzy-based CREAM approach is established for the evaluation of human performance reliability in dynamic changing situations. A new list of common performance conditions (CPCs) depicting the main features of situated driving context is defined. Three data clustering approaches to determine the membership functions are explained. The new concept of human performance reliability score (HPRS) to quantitatively evaluate human performance is established. In chapter 5, situated driving context is taken as an example to explain the modified fuzzy-based CREAM approch. The experiment results are analyzed including the unfuzzified HPRS results and HPRS results from data clustering approaches. The monitoring and evaluation of human driver reliability in situated context is explained. The summary and outlook of this work is given in chapter 6.

2 Human behaviors and reliability approaches

In this chapter, human reliability related terms are collected and explained. The wildly used 'three generation' of human reliability analysis approaches are reviewed. The research gaps for the existing HRA methods are presented.

Part of the contents, figures, and tables presented in this chapter are modified after previous publications [HS22][HLLS21][HLL21]. Part of the contents, figures, and tables are prepared for publication of [He22b][He22a].

2.1 Human reliability-related concept

2.1.1 Human error definitions

Two views are distinguished between 'old view' and 'new view' of human error [Dek17]. In the 'old view', human error is the cause of trouble and it is a simple problem, when all systems are working well, people just need to pay attention and comply to avoid human errors. People can, and must, achieve zero errors, zero injuries, and zero accidents. In the 'new view', human error is a symptom of deeper trouble and the complexity of generating human error is depending on the complexity of the organization and environment. People can, and must, enhance the resilience of the people and organization. It could be concluded that the understanding regarding human errors gradually shifts from asking who is responsible for the outcomes to finding out what is responsible for the outcomes.

In the process of understanding human error, different definitions and related glossary of terms have been proposed. Swan and Guttman defined human error as an error that is simply an action which is out of tolerance, where the limits of the tolerance is defined by the system [SG83]. From Rasmussen's point of view [Ras82], human error can only be described with reference to human objectives or expectations, it depends on the explicit situation. From Reason [Rea90], it is obtained that human error is taken as a universal term to comprise all the occasions which a planned sequence of mental or physical activities fails to generate the intended outcome, and these failures cannot be associated to the intervention of some chance agency. Hollnagel defined human error as an erroneous action which fails to generate the expected result and/or which produces an unwanted consequence [Hol98]. In Dhillon's definition [Dhi17], human error is the failure to execute a stated task that could result in interruption of scheduled operations or damage to property and equipment.

2.1.2 Human error taxonomies

Various human error taxonomies have been proposed. Three dominated taxonomies are reviewed in this chapter, which are Rasmussen's skill, rule, and knowledge error [Ras87a], Reason's slips, lapses, mistakes and violations [Rea90], and Hollnagel's phenotypes and genotypes [Hol98].

In Rasmussen's skill-,rule- and knowledge-based (SRK) behavior model, errors are affected by skills, experience and familiarity with the situation encountered. The generic error-modeling system (GEMS) is applied to classify these errors [Rea90].

- Skill-based behavior is developed without conscious control as smooth, automated, and highly integrated patterns. Skill-based error is typical detected in routine repetitive work.
- In rule-based behavior, the actions are often controlled by a memory-based stored rule or procedure.
- The performance which is goal-controlled during unfamiliar situations, which no rules for control are available is knowledge-based behavior.

Reason classified human errors into slips, lapses, mistakes and violations. When combining with Rasmussen's SRK model, skill-based errors correspond to slips and lapses, rule-based and knowledge-based errors are related to mistakes.

- Slips are errors which result from some failures in the execution of an action sequence. Slips can be seen as externalized actions not conducting as planned.
- Lapses are errors which result from failures in the storage stage of an action sequence. Lapses are generally used for more covert error forms, including failures of memory.
- Mistakes are failures in the inferential and/or judgemental processes in the selection of an objective. Mistakes are more subtle than slips and harder to detect.
- Violations relate to actions habitual or isolated departure from rules and regulations.

In Hollnagel's CREAM approach, it is stated that human actions/errors are all to some extent cognitive, indicating that they are not able to be properly described without consideration of human cognition. Human error can be identified as phenotypes and genotypes.

- Phenotype concerns with the manifestation of an erroneous action. It can be divided into action at wrong time, action of wrong type, action at wrong object and action in wrong place/sequence.
- Genotype refers to the possible causes such as the functional characteristics of the human cognitive system that are assumed to contribute to an erroneous action. Human related genotypes can be further divided into observation, planning, interpretation, temporary person related causes and permanent person related causes.

2.1.3 Performance shaping factors (PSFs)

Human error in human reliability analysis (HRA) is not viewed as the product of individual shortcomings but rather as the combined impact of contextual and situational factors on human performance [Bor12]. These factors are denoted as performance shaping factors. A PSF can be anything affecting the ability of an individual to complete a task [Phi18]. It is important to understand PSFs so to quantify human error probability as PSFs may lead to human errors.

In general, PSFs are classified into internal and external, corresponding to the individual and situational or environmental circumstances, respectively. In [BGJ07], the new classification of direct and indirect for PSFs are trying to be established based on the relationship between magnitude of the PSF and how the magnitude is measured. But the listed PSFs are just examples to show how direct and indirect PSFs are classified. In [Phi18], a detailed PSFs list including internal and external PSFs are listed. For the further categorization, internal PSFs can be divided into three groups: temporary physical and mental states, permanent physical and mental states, and readiness for duty. There are 104 internal PSFs and 20 external PSFs collected. With PSFs considered in human performance evaluation, the quantification of human error probability becomes possible.

As the PSFs vary widely with the operation characteristics in different application fields, when discussing the main PSFs, they are mainly discussed separately according to different application fields. For example, in [KMO15], the most significant performance shaping factors in railway operations are identified with the analysis of 479 incidents and accidents in railway operations, which are safety culture, system design, fatigue, communication, distraction, quality of procedures, perception, training, expectation, quality of information, supervision, and workload. The number of PSFs necessary for human reliability analysis is not identical in different HRA methods, ranging from single factor model such as time-reliability curves up to 50 or more PSFs in some new developed HRA models [Bor10]. In different HRA approaches, PSFs is called by other terminologies. For example, PSFs is called error producing conditions (EPCs) in human error assessment and reduction method (HEART) [Wil88], and common performance conditions (CPCs) in cognitive reliability and error analysis method (CREAM) [Hol98].

2.1.4 Human error probability (HEP)

Human error probability (HEP) is a variable to characterize the probability of human error occurrence or briefly: the reliability of humans [DPMIR15]. The definition of HEP could be summarized as the mathematical ratio between the number of errors occurring in a task and the number of tasks carried out with the opportunity for errors. The number of opportunities for error is generally the same as the number of times the task is carried out [Whi04]. The mathematical expression can be simply expressed as

$$HEP = \frac{number\ of\ observed\ errors}{number\ of\ opportunities\ for\ errors}. \tag{2.1}$$

The HEP is the indicator for the relative occurrence of errors and subsequently faultless actions. In Table 2.1, how HEPs are calculated in some representative HRA methods is shown. It can be obtained that the final HEP is deeply affected by PSFs which are highlighting human error contributors and adjust basic human error probabilities. In general, experience, complexity, stress, adequacy of procedure, human-system interface, and workload are adopted as PSFs in HRA [PJK20].

In [BR16], the HEP with the consideration of PSFs evolution and progression of events results in the dynamic HEP varying with time, which is compared with the traditional or static HEP with non-effect of time on the error estimation. For dynamic HEP, the equation could be expressed as

$$HEP_{dynamic} = f(HEP_{nominal}|PSF(t)), \tag{2.2}$$

where t is time. The equation of static HEP can be represented as

$$HEP_{static} = f(HEP_{nominal}|PSF). \tag{2.3}$$

It could be detected that the difference between Equation 2.1 and Equation 2.2 is the consideration of PSFs evolution with time. Figure 2.1 and Figure 2.2 draw the dynamic HEP and static HEP with time, respectively. It is obtained that the values of dynamic HEP vary with time while the values of static HEP are constant.

2.2 Human reliability analysis approaches

2.2.1 Overview of human reliability analysis

Human reliability analysis (HRA) is a structured methodology applying qualitative and quantitative methods to assess human contributions to system risk, which consists of the determination the effects of human errors to the system, prediction of

Table 2.1: Equations for HEPs in various HRA methods [He22b]

HRA methods	Equations for HEP
THERP [SG83]	$HEP_{Final} = BHEP_{Diagnosis} \cdot \prod_1^n PSF_{Diagnosis,i} + BHEP_{Execution} \cdot \prod_1^n PSF_{Execution,i}$
SLIM-MAUD [E+84]	$SLI = \sum(NormalizedWeight(PSF_i) \cdot State(PSF_i))$; $Log(1-HEP) = a \cdot SLI + b$
SPAR-H [GBM+05]	$HEP = NHEP \cdot \prod_1^8 S_i$; $HEP = \frac{NHEP \cdot \prod_1^8 S_i}{NHEP \cdot (\prod_1^8 S_i - 1) + 1}$
HEART [Wil88]	HEP =Nominal human unreliability$\cdot \prod$Assessed effect$_i$
ATHEANA [RSN]	$P(HFE\|S) = \sum_j \sum_{i(j)} P(EFC_i\|S) \times P(UA_j\|EFC_i, S)$
HuRECA [LKJ12]	$HEP_{diag} = BasicHEP_{diag} x \prod w_i(PSF_i)$; $HEP_{exec} = \sum[BasicHEP_{exec}(i) \times HEP_{rec}(i)]$

THERP: Technique for human error rate prediction; BHEP:Basic HEP.
SLIM-MAUD: Success likelihood index method using multi-attribute utility decomposition; a and b: Constants that can be obtained by two sets of known HEPs.
SPAR-H: Standardized plant analysis risk HRA; S_i: The multiplier associated with the value of corresponding PSF levels; NHEP: Normal HEP, for diagnosis task is 0.01 and for action task is 0.001.
HEART: Human error assessment & reduction techique; Assessed effect=((Multiplier of EPC-1) $\times$ Assessed proportion of effect)+1; EPC: Error producing conditions.
ATHEANA: A technique for human event analysis; $P(HFE|S)$: Probability of the error for the HFE applicable to accident scenario (S); $P(EFC_i|S)$: Probability of accident contexts to the scene of accident including deviations and nominal context; $P(UA_j|EFC_i, S)$: Probability of failure of UA corresponding to each context evaluated; HFE: Human failure event; EFC: Error forcing context.
HuRECA: Human reliability evaluator for control room actions; $Basic_H EP_{diag}$: $f(availabletimefordiagnosis)$; $Basic_H EP_{exec}(i)$: $f(tasktype(i), stresslevel(i))$; $HEP_{rec}(i)$: $f(availabletime(i), MMI(i), supervisorrecovery(i))$.

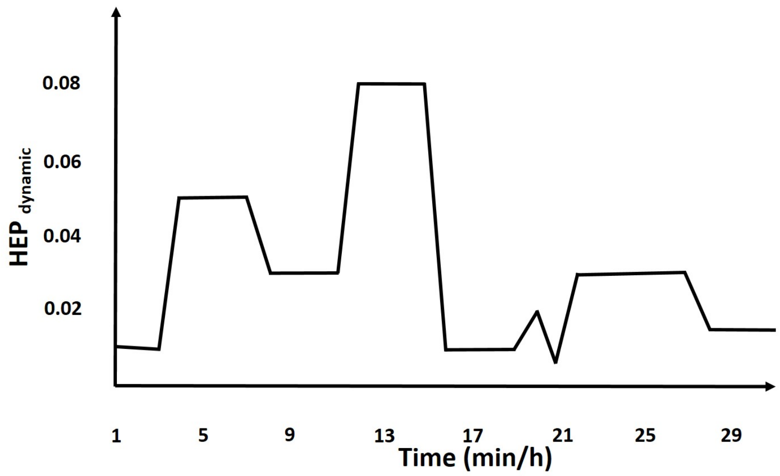

Figure 2.1: Effect of time on the error estimation in dynamic HRA (adapted from [BR16])

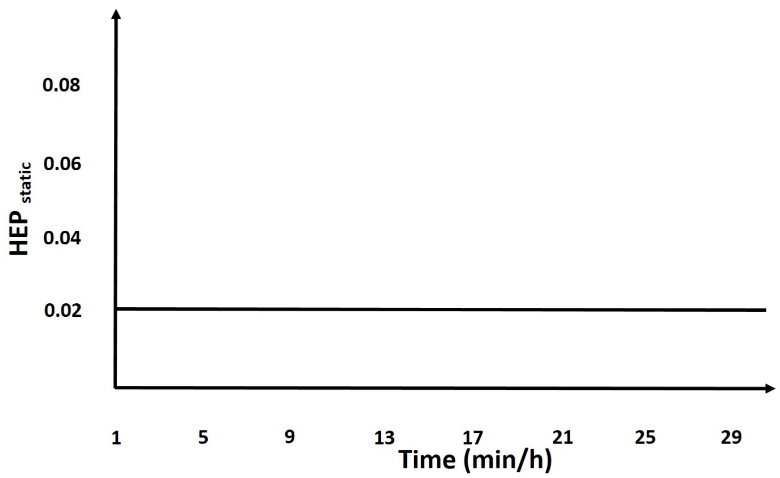

Figure 2.2: Non-effect of time on the error estimation in static HRA (adapted from [BR16])

the probability of the error occurrences, and the identification of the potential consequences [Phi18]. The HRA methods are initially developed for the evaluation of human performance in nuclear power industry. With the development of technology, reliability of hardware and software in systems are improving continuously, human performance reliability is a growing concern as the proportion of human-factor related accidents is increasing. Therefore, HRA methods are applied to other safety critical industries with well-developed procedures. The general steps for HRA is presented in Figure 2.3.

In HRA process [Phi18], problem definition is to determine the scope and type of

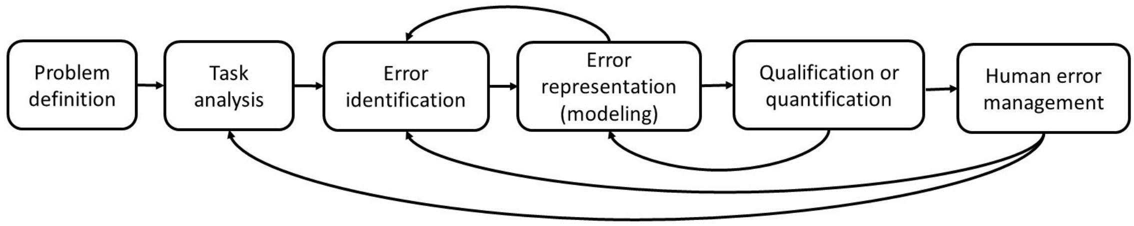

Figure 2.3: Basic steps in the HRA process (adapted from [Phi18])

analysis, the tasks that will be evaluated, and the human actions which need to be assessed. Task analysis is also known as task decomposition, which includes the identification and breaking down of each task into the steps and sub-steps systematically that constitute the human activities necessary to achieve the task goal. The task can be decomposed into different levels depending on the purpose of the task analysis and the information available. Human error identification is the most important step in HRA as failing to identify a human error may lead to the omission of the contribution of the error to risk and eventually to the underestimation of the whole system risk. In this step, not only the types of human error that could occur need to be identified, but also the factors that contribute to the occurrence of error should be determined. Human error representation or modeling is used for better understanding the the causes, vulnerabilities, recoveries of human errors. The tools could be used for human error representation contain master event trees, fault trees, and event sequence diagrams, etc. Human error quantification is to assign probabilities to human errors with the calculated human error probability (HEP). The final step in HRA is human error management which includes the establishment of barriers to prevent errors, providing means to detect and correct errors, and modifying the performance shaping factors that negatively affect human performance.

To accomplish the HRA process, many structured methods are generated and with the development, these methods can be briefly divided into three generations where the 'first generation' of HRA methods concern with the quantification of human errors to calculate the HEP of tasks, the 'second generation' of HRA methods more focus on the human cognitive process during the task for the qualitative explanation of human reliability, and the 'third generation' of HRA methods consider the dynamic features in human performance.

2.2.2 The 'first generation' of HRA

For the so called 'first generation' HRA methods, human is considered similar to a mechanical component, so all aspects of dynamical interactions with the working environment, both physical and social environment are not fully considered [DPIMR13]. The basic assumption which has been made in many of these methods such as technique for human error rate prediction (THERP) [KKTAL97], accident sequence evaluation program (ASEP) [Swa87] and human cognition reliability (HCR) [HSL85] is that humans have natural weakness and logically fail to execute tasks, similar to mechanical or electrical components. With this assumption, based on the operator's task characteristics, the HEP can be assigned by experts and can be modified by performance shaping factors (PSFs). In the 'first generation' of HRA methods, the characteristics of the task are considered as main factors, which could be represented by HEPs; the context, represented by PSFs, is regarded as the minor factor in evaluating human performance reliability [KSH06]. The main features of the 'first generation' of HRA methods could be concluded as follows [Kim01]:

- Human reliability is similar with hardware/equipment reliability in conventional reliability analysis.
- Human actions either succeed or fail to carry out a given task.
- Human errors are distinguished into two categories, one is failure to perform an action known as an 'omission' and the other one is an unintended or unplanned action known as a 'commission'.
- The phenomenological aspects of human actions are focused on, which means that an operator could either do something correctly, do it incorrectly (i.e. commission), or not do it at all (i.e. omission).
- Cognitive aspects of human actions are less concerned in the so called 'first generation' HRA methods although some cognitive models are adopted in some of these methods.
- The quantification of human errors is emphasized to calculate the human error probabilities (HEPs).
- The context is indirectly treated. The influence of PSFs (i.e. context) on the operator performance is simply taken into account by multiplying the nominal HEPs with a weighted sum of PSFs.

Example: THERP

The most popular and effective method among the 'first generation' techniques is technique for human error rate prediction (THERP), which characterized by an accurate mathematical treatment of the probability and error rates as well as well-structured fault trees for the evaluation of human error. The framework of THERP is event tree modeling with each limb representing a combination of human activities, the effects and results of human activities [GM11], as shown in Figure 2.4. The nodes in the tree indicate actions, the sequence of actions in the task is presented from top to bottom. Two branches are originated from each node, the branches to the left, marked with lowercase letters, indicate action success; the branches to the right, marked with capital letters, indicate action failure. The action failure probability of each action are denoted as F_A, F_B, and F_C.

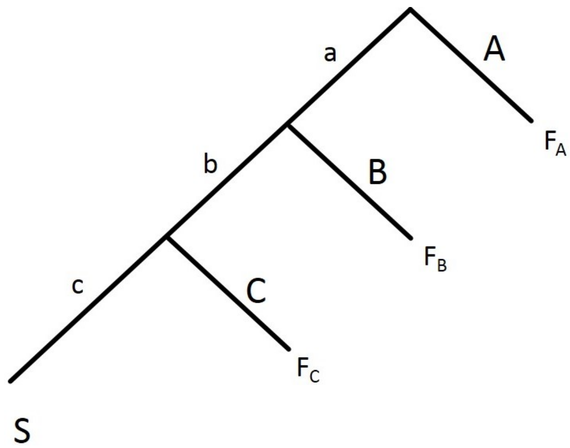

Figure 2.4: Framework of THERP event tree (adapted from [DPIMR13])

The HEPs in THERP are calculated through three steps which are analyzing event, quantifying effects of factors and interactions, and calculating human error contribution to probability of system failure [CCM+06]. In the first step, the HEP search scheme is applied for each branching point of the event tree to identify the possible human errors and the related basic/nominal HEPs. Therefore, the factors and synergy which affect human performance are identified. In step two, the levels of task dependencies are assessed based on the five-level dependency scale which is specified by THERP. The possibilities of recovery from errors are accounted for, the possible recovery branches in the event tree and success probabilities are assessed. In the last step, the success and failure consequences are analyzed within the event tree, the HEP is calculated and applied to the system model. The HEP can be expressed as follows [RNR+20]

$$HEP = BHEP \times PSF_1 \times PSF_2 \times \cdots \times PSF_n, \tag{2.4}$$

where BHEP indicates basic HEP given in the related THERP tables.

2.2.3 The 'second generation' of HRA

With the criticism of absence of consideration on the dynamic aspects from the environment, researchers developed some new methods, such as a technique for human event analysis (ATHEANA) [CRSWP96], and cognitive reliability and error analysis method (CREAM) [Hol98]. These methods are so called 'second generation' HRA methods. The methods in this generation aspire to lean toward conceptual methods as cognitive models are proposed, while the so called 'first generation' methods are often behavioral approaches. The qualitative assessment of the operator's behavior and the research for models which describe the synergy with the production process are the aim of the 'second generation' HRA methods [DPIMR13]. The focus of the 'second generation' HRA methods shifted to the cognitive aspects of humans, the causes of errors rather than their frequency. Cognitive models are developed to explain the logical-rational process of the operator, the dependence on personal human factors, and the interaction between humans and external environment. In this case, operator should be a complicated and integrated system to collaborate to achieve the task.

The 'second generation' HRA methods are more appropriate to explain human behavior as these models are based on cognitive models. Cognitive models are an essential tool for understanding human performance. The immediate solution to apply human cognition in HRA methods is to introduce a new category of error, namely, 'cognitive error'. It defined not only as failure to accomplish an activity which is mainly of a cognitive nature, but also as the cause of activity that fails [Hol98].

The main features of the 'second generation' of HRA methods could be concluded as follows:

- Human cognition models are used to identify why errors happen.
- Context is the most important factor affecting human reliability.
- The majority of the proposed approaches rely on implicit functions relating PSFs to probabilities.
- The approaches have yet to be empirically validated.
- It is lack of empirical data for model development and validation.
- It is lack of inclusion of human cognition (i.e. The methods need for better human behavior modelling).
- The variables in 'second generation' HRA deeply depend on the methods used.
- The evaluation of human reliability with 'second generation' heavily rely on expert knowledge in selecting PSFs and use of these PSFs to obtain the HEP.

Example: ATHEANA

The ATHEANA approach is a widely recognized 'second generation' HRA method to model human behaviors through a cognitive model of information processing, including monitoring, diagnosis, planning, and execution [PMS15]. From [Kim01], the approach started with the human failure events (HFEs) identification from the accident scenarios of PSA model. Next, unsafe actions (UAs), indicating actions inappropriately taken or not taken when needed in a degraded plant safety condition, are introduced to represent the HFEs. Furthermore, the error forcing context (EFC) is defined which is the combined effect of performance shaping factors (PSFs) and plant conditions. At the end, all factors are are taken into account for the HFE probabilities estimation. The framework of ATHEANA approach is presented in Figure 2.5.

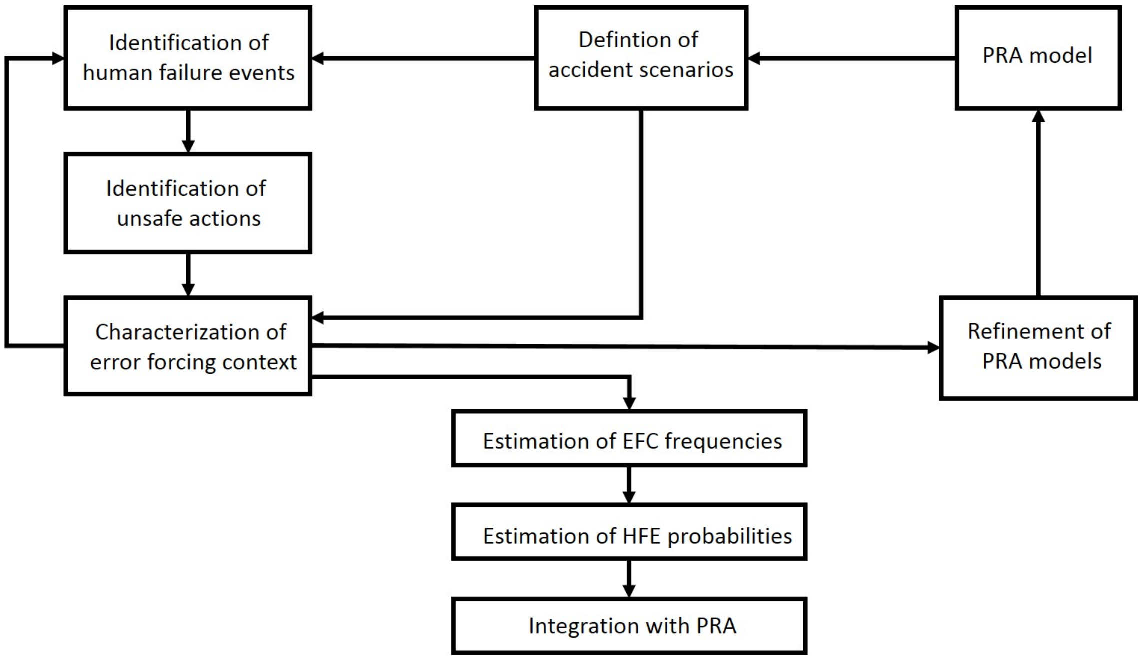

Figure 2.5: ATHEANA HRA approach (adapted from [Kim01])

The HFE probability in scenario S is the conjunction of probabilities of EFC (i is the number of forced errors) and UA (j is the number of unsafe actions) in a dependable form as [PMS15]

$$P(HFE|S) = \sum j \sum i(j) P(EFC_i|S) \times P(UA_j|EFC_i, S). \tag{2.5}$$

The characteristics of ATHEANA approach could be concluded as follows [Kim01]:

- It offers very detailed instructions on how to determine the cause of the HFE-related unsafe actions, i.e. the error-forcing context.

- The ATHEANA method with HFEs from PSA model contains many drawbacks that the consequences of human errors are limited by the pre-identified PSA accident sequences.
- The theoretical background of ATHEANA seems weak for predictive analysis as the cognitive model are useful for post accident human reliability analysis and post accident human error actions are often cuased by cognitive error.

2.2.4 The 'third generation' of HRA

The development and characteristics of the 'third generation' of HRA are discussed. A representative method is illustrated as an example.

Further improvements for existing methods are driven by the limitations and deficiencies of the 'second generation' HRA methods. There are also studies that focused on the shortage of empirical data for the development and validation of an HRA model and are intended to define the database HRA, which may provide the methodological tools needed to more intensively use types of information in future HRA methods and reduce uncertainties in the information used to conduct human reliability assessments [DPIMR13]. With the increased development of computer technology, several HRA methods are using artificial intelligence and simulation techniques to predict human error based on cognitive models. The cognitive simulation model (COSIMO) [CDD+92] based methods are defined as so called 'third generation' HRA methods [PLH17]. From [GSM19], the requirements for the 'third generation' HRA methods must include: i) comprehensive (addressing a joint system of humans and machine, providing explicit representations of the causal factors for human-machine failures with cognitive science and system engineering, addressing the full spectrum of contexts and causal factors related to HRA), ii) research-based (connecting with multiple sources, types, and sizes of data, models, and information), iii) adaptable and flexible (accommodating changes in database structures, data sources, and methodologies), and iv) multi-purpose (quantitative and qualitative aspects of HRA). The nuclear action reliability assessment (NARA) is now defined as 'third generation' HRA method [DPIMR13].

The so-called 'third generation' HRA methods concern with the reliability dynamics and simulation and modeling of human performance are adopted. The use of simulation and modeling in HRA to capture and generate data is presented in Figure 2.6.

Example: NARA

The NARA approach, as a data-based HRA tool, is developed to concentrate on the nuclear context as the widely used human error assessment and reduction technique

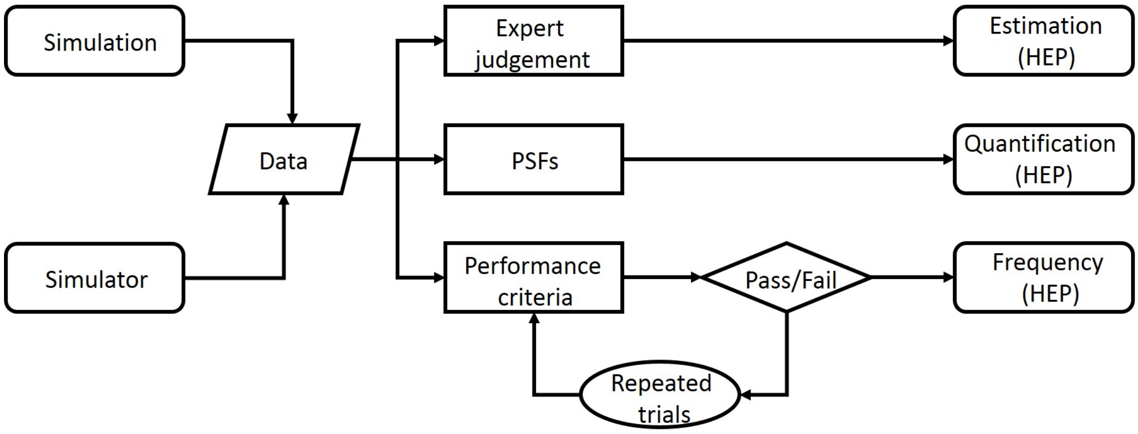

Figure 2.6: Uses of simulation and modeling in HRA (adapted from [Bor07])

(HEART) did not always fit very well with the nuclear power plant tasks [GKK+17]. Generic task types and corresponding human error probabilities are employed as one part of the quantification process. The key elements of NARA are [GKK+17] listed below.

- Generic task types (GTTs). A task needs to be quantified is assigned a nominal human reliability by classifying it into the generic task types.
- Error producing conditions (EPCs). The factors negatively affect human performance are identified based on a set of EPCs with comparing to the 'ideal' conditions associated with GTTs.
- Assessed proportion of affect (APOA). The analyst judges the strength of the affect that the EPC has on the task performance.

The new features in NARA comparing with HEART in this specific case are [BH09]:

- Quantifying operator reliability for long time-scale events.
- A prototype concept to error of commission quantification.
- More instructions have been developed for use of APOA process.

2.3 Research gaps

The first gap in the existing HRA methods is the lack of data for model development and validation. In the existing 'first generation' and 'second generation' of HRA methods, the empirical data from HRA experts are used to define human reliability as a scalar value. In the so called 'third generation' HRA methods that are still in development, the gap of shortage of empirical data is being tried to be covered [DPIMR13]. Some methods are established with artificial intelligence and simulation techniques to predict human error by computer-based modeling and simulations [PLH17]. However, these methods are mainly applied to crews in industrial factories and nuclear power plants. When considering the human reliability of the individual, the existing HRA methods cannot be adopted properly.

The second gap within the existing HRA methods is the missing consideration of human reliability in situated and therefore dynamic context. The existing 'first generation' and 'second generation' HRA methods are considered as 'static' HRA methods as events are analyzed for an assumed window of time [BR16]. The window of time ranges therefore from different failure event and allows a rough specification of relations. The human error probability for 'static' HRA is not changing as a function of time or the event progression as the PSFs are not considered to evolve over time. For the dynamic human reliability, the evolution of PSFs with time and their consequences to the outcome of events should be accounted for. Although the so called 'third generation' of HRA methods consider the dynamic progression of human behavior, this progression is measured on event-based level, rather than on action-based level [BR16]. As a result, the HEP results of dynamic HRA on event-based level may bounce between discrete values, while the HEP results on action/scene-based level fluctuate in continuous values as the time of action/scene-based HEP is measured in second scale and the time of event-based HEP is measured in minute or hour scale depending on the number of sub-tasks or actions containing in the event. As a conclusion about the actual state-of-the-art in this field, it can be stated that the existing HRA methods are not suitable for the evaluation of individual human reliability in situated and dynamic context.

The third gap in the existing HRA methods is the heavily reliance on expert knowledge in selecting the PSFs, determining the behavior levels, and calculating the HEP. For example, in the often used 'first generation' HRA method of THERP (technique for human error rate prediction), the effects of human errors on the system failure events should be estimated by expert knowledge to determine which HEP should be selected considering the predefined nominal HEP and error factors [SG83]. In the well-known 'second generation' HRA method of CREAM (cognitive reliability and error analysis method), the common performance conditions need to be assessed by expert knowledge to define the behavior levels which correspond to the effects on performance reliability [Hol98].

3 Quantification of human behavior levels

In this chapter, the cognition models is reviewed, the SRK model is the focus and explained, how to quantify human behavior levels in SRK model is presented. The relationship between human behavior levels and HEP are determined, and the effects of time pressure and training on SRK levels switching are analyzed. Based on this, a more general framework on the SRK levels switching are proposed and the expected applications are discussed.

Part of the contents, figures, and tables are prepared for publication of [He22b].

3.1 Cognition model

3.1.1 Cognition concept

Cognition is related to human behaviors of knowing, perceiving, and thinking. In [Hol98], it is argued that for technically minded people, cognition could be explained in loose term as which went on in the head. Cognition models are developed to explain human information processing. The information obtained from sensory systems is put into the mind or 'black box' of the human to process and a response is generated [Wha16]. The 'black box' is used as the cognitive processing is not able to be seen. In this case, many mechanisms are development to explain the process of human cognition, the well-known models are the IDAC (Information decision and action crew) [MC04], human information-processing model [WHHB21], SRK (skill, rule, and knowledge) framework [Ras83], and situation awareness mechanisms [End17].

Despite various cognition models are developed and different terms are adopted, the fundamental functions used in these models are more or less similar, which include detecting, understanding, decision making, and action. In [Wha16], these terms are explained.

- Detecting. Detecting is the process of perceiving information from context and focusing selectively on information that is relevant to present activities. The cognitive processes related to detecting are sensation, perception, and attention.
- Understanding. Understanding is to understand the meaning of the information that has been detected. The cognitive processes included in understanding are sense making, situation awareness (SA), interpretation of the information, and integration of information together for diagnosis.

- Decision making. The function of decision making contains goal selection, planning, re-planning and adapting, evaluation options, and selection.

- Action. The definition of action is related to the implementation of an action on the level of a single manual action (for example operating a valve) or a predetermined sequence of manual action. The action contains the alteration of plant status as a result of manipulating hardware and/or software.

3.1.2 Cognitive process

The HRA methods use different cognitive explanatory models of the human behavioral process to explain the mechanisms by which human errors occur. The cognition is mainly related to judgment from memory, interpretation, concept formation, decision making, and other mental activities before action execution in the environment. A cognitive model is generated to describe human cognitive process and explain human thinking and behavioral modes. With the development on psychology, behavioral science, ergonomics, and other interdisciplinary, the understanding on human cognition becomes more detailed. At the same time, these cognitive models inspire HRA researchers to develop more comprehensive HRA methods on different human cognitive activities. The HRA methods with their adopted cognitive processes are listed in Table 3.1.

Among the listed cognition models, skill-rule-knowledge (SRK) model proposed by Rasmussen is well-known and wildly used. This model has been applied in many application fields in human-machine system [She17]. In 1979, Rasmussen was able to distinguish human behavior into three levels including skill-based behavior, rule-based behavior, and knowledge-based behavior [Woo09]. It is known that skill-based behavior corresponds to highest human reliability and knowledge-based behavior has the lowest human reliability from the consideration of cognition process, but this is only considered qualitatively [LC15]. Although human error probability (HEP) intervals of SRK model are estimated in [GB93], and modified in [SXSL09], the data are only taken within a THERP context. No quantitative results on the human reliability of these three different levels of behaviors using data from different generation of HRA methods exists in existing research. However, this is of increasing importance as automation in human-machine systems is becoming increasingly important. Human skills are changing from a fundamentally technical understanding of devices to abstract process management skills. The question arises whether humans can control automation in certain challenging situations (takeover situations or when the driving state abruptly changes), whether autonomous driving vehicles make human qualification for vehicle guidance superfluous? In addition to issues of disqualification [VD20] and training [DDWN09], the question arises in practice about the right time for the warning [TS19] or for suitable interfaces [WS18].

Table 3.1: Human cognitive process and corresponding HRA methods (adapted from [PLH17])

Researcher	Year	Cognitive process	HRA methods
Rasmussen	1979	Recognition, identification, decision, and planning	THERP/HCR
Wreathall	1982	Detect, diagnose, and implement	OAT
Woods	1987	Monitoring, explanation building, and response management	CREATE
Reason	1990	Planning, storage, and execution	GEMS
Cacciabue	1992	Cognitive filter, diagnosis, hypothesis evaluation, and execution	COSIMO/DYLAM/HERMES
Wickens	1992	Perception, decision and response selection, and response execution	ATHEANA/CREAM/ADS-IDAC
Cooper	1996	Monitoring, situation assessment, and response	ATHEANA
Hollnagel	1998	Observation, interpretation, planning, and execution	CREAM
Mosleh	2004	Information processing, problem solving and decision making, and execution	ADS-IDAC

ATHEANA: A technique for human event analysis.
ADS-IDAC: Accident dynamics simulator with the information decision and action in a crew context operator model.
CREAM: Cognitive reliability and error analysis method.
COSIMO: Cognitive simulation model
CREATE: Cognitive reliability assessment technique.
DYLAM: Dynamic logic analysis method.
GEMS: Generic error modeling system.
HERMES: Human error reliability methods for event sequences.
HCR: Human cognition reliability.
OAT: Operator action tree.
THERP: Technique for human error rate prediction.

To answer these questions, the human reliability must be quantitatively evaluated to generate knowledge about the quantitative knowledge about the human reliability within the context of new relations between human and machine. The goals of this chapter include:

1) Quantification of human behavior levels in SRK model;
2) Analysis of the effects of time pressure and training on SRK levels switching;
3) Establishment of a general framework to map the relation between HEP and SRK levels.

To determine HEP of these three levels, three databases (technique for human error rate prediction [KKTAL97], Savannah river site human reliability analysis (SRS-HRA) [BHO+94], and nuclear action reliability assessment [KGK+04]) as HRA methods are used.

3.2 Skill-rule-knowledge (SRK) framework

3.2.1 The SRK level behaviors

According to Rasmussen's study [Ras82], human behavior can be differentiated into categories according to different ways of representing the restrains in the behavior of a deterministic environment or system, three different kind of interaction with respect to the integration of human cognitive abilities can be distinguished with related different performance results: skill-, rule-, knowledge-based performance. Whether or not the operator is involved in problem solving at the time an error occurred is the key distinction based on SRK levels [Rea90]. These levels and a brief illustration of their relations are shown in Figure. 3.1

Skill-based behaviors

According to [WHLS14] the sensory-motor performance along with a statement of intentional acts or activities, development with noncognitive control as smooth, automated, and highly integrated patterns of behavior is an indication of skill-based behavior. Those actions have been rehearsed in a more or less training process and then go by in a steady flow. Such well-established skills are the most effective forms of human behavior in terms of time and strain. Skill-based behavior is typical for routine repetitive work, even leaving room for secondary activities that may not necessarily be task-related.

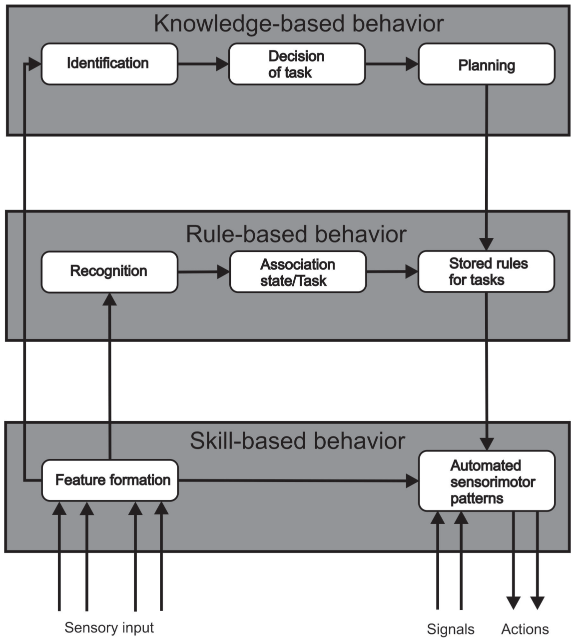

Figure 3.1: Rasmussen's SRK (skill-rule-knowledge) model (adapted from [Ras82])

Rule-based behaviors

In the rule-based behaviors, the sequence architecture of subroutines in a well-known work situation is consciously controlled by a memory-based stored rule or procedure which may have been derived empirically during previous experience, communicated from other individuals' know-how as an instruction or a handbook.

From [Ras87b], the border between skill-based and rule-based behaviors is ambiguous, and depends on the level of training and on individual attention. In general, the skill-based behaviors perform without conscious attention. The individual will be unable to explain how to control the performance and is unable to explain the information based on for the performance. Explicit know-how is referred to in a higher level of rule-based coordination. The rules used can be reported.

Knowledge-based behaviors

The performance which is goal-controlled during unfamiliar situations, which no know-how or rules for control are available from previous encounters, is known as knowledge-based level. This type of behavior can be described as a mental process in which the operator searches for problem-solving action options based on knowledge that is already known or that has still to be learned. In doing so, the operator checks whether the thought-out action routines are suitable for the goal-oriented management of the situation and finally applies the solution that seems to be the most effective. The process is thus highly individual and is always based on the existing knowledge and cognitive abilities [Söf01b]. Otherwise, there will be no action usually because there is not enough time [WHLS14]. Successful solutions may be stored as rules for future challenges.

From the description of these three behavior levels, it can be found that these three levels could be interacted in between. The interaction is useful when a task performance needs to be analyzed, because the cognitive activities are not always at the same level, but will shift to another. When a disturbance occurs acting on the skilled performance from the environment, the attentive cognitive apparatus searches for suitable rules to adjust the performance. At knowledge-based domain, the attention is usually planning for future activities or improvement of rules from previous successful applications.

3.2.2 Interaction between levels

In this thesis, a dynamic system is defined as the outside world of the interacting human in which the relevant variables are dynamic and therefore enforce a dynamic human interaction. The dynamic environment will result in the dynamic interaction, which will further include that the time variant changes are perceived by human cognition. It is known that human cognition is not only strictly linear or serial, but also involves parallel and cyclic processing [Wha16]. Cognitive functions occur in a continuous loop and overlap. Operators in the realistic context often need to accomplish most or all of these functions at the same time [KRM+03]. Therefore, task performance normally require a simultaneous consideration of all three cognitive control levels in SRK model. Each level of cognitive control may be focused on different aspects of the task at a given time, and several activities may be ongoing at the same time [RV89]. The interaction of human behaviors between different levels in SRK framework is shown in Figure 3.2.

From Figure 3.2, the conscious attention is free to deal with other activities on a time sharing activities during skill-based routines execution, which is shown as synchronous activities. The rule-based domain is involved in retrieving an appropriate rule-set from the memory if the next task requires an activity sequence which are not integrated into an automated pattern. Interruptions may occur during the task performance when choices are needed to be made or adjustments of the current activities are needed, which is presented as synchronic activities. Offline planning consideration may be neither happened in the same time frame, nor in the same part of problem space. Attention may be switched to the evaluation of past activities and planning for the further activities when skilled activities are processing smoothly, which needs knowledge-based analysis and planning, shown as achronic activities, or the recall and evaluation of the success rules from previous performance, presented as diachronic activities.

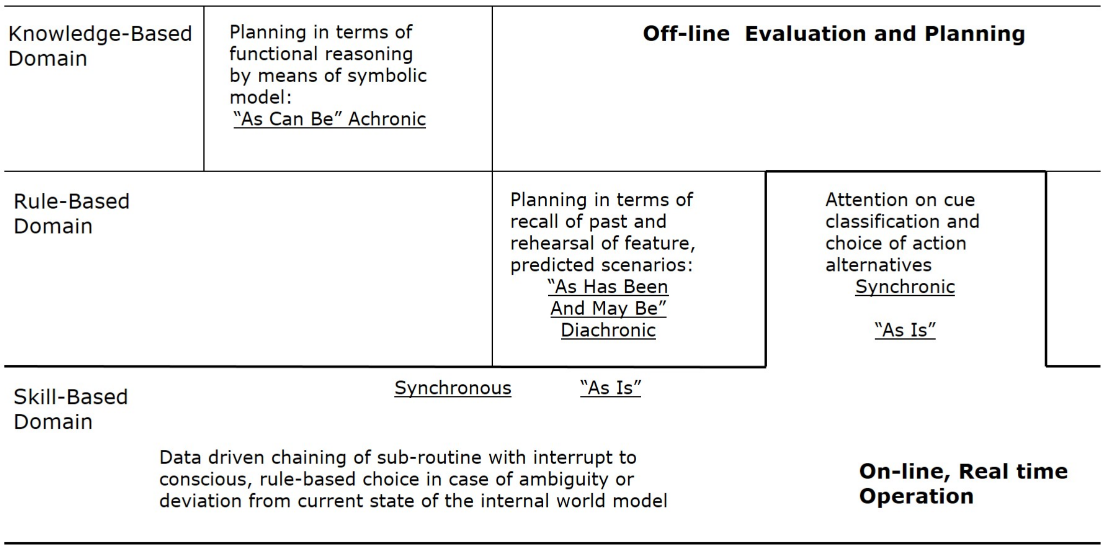

Figure 3.2: Interaction of human behaviors between different levels (adapted from [RV89])

3.3 Quantification of human behavior levels with SRK model

3.3.1 Databases

Three databases, including technique for human error rate prediction (THERP) from so called "first generation", Savannah river site HRA (SRS-HRA) and nuclear action reliability assessment (NARA) from so called "third generation", are selected to obtain operation tasks and corresponding HEPs. These three databases are mainly applied in the nuclear power plant (NPP) field. The data are generated from surveys on skilled operators, advice from HRA experts, and site visits. In recent years, SACADA (scenario authoring, characterization, and debriefing application) [ZSG17] and HuREX (human reliability data extraction) [JPK+20] are often referred within the NPP field. In this study, the three databases are selected which are based on a broad and widely representation of application fields.

The database of THERP is based on studies and observation from various kind of plants in the world. Besides that, it also obtained support and guidance from program managers at nuclear regulatory commission [SG83]. In THERP, 44 tasks are analyzed. For each task operation, a basic HEP and corresponding error factor (EF) exist, where the basic HEP denotes the probability of human error without considering the conditional influence of other tasks [SG83]. The error factor is integrated for the variation in estimated HEP due to different operation conditions and modeling uncertainty. For operation at different conditions or environments, HEP will be various. The upper bound of the estimated HEP is the product of basic HEP and EF, while the lower bound of the estimated HEP is the result of dividing HEP by EF.

The SRS-HRA database is developed from generic models and SRS-specific data, surveys from department of energy sites, THERP, human cognitive reliability (HCR), and actual national or regional data for transportation accidents and expert judgment [BHO+94]. In this database, 35 human error events with 3 different failure probabilities: low, nominal and high are considered. The nominal or low HEP is chosen for a situation with normal operation, planned process transients, and frequent minor abnormal occurrences. Nominal or high HEP is selected when the situation is less frequent, more significant abnormal occurrence. High HEP is applied when the effects are directed on personnel (e.g., personal well-being threatened).

The database from NARA comes from computerized operator reliability and error data (COREDATA) which are supported by a wide range of information, thus understanding of HEP in its practical and methodological context becomes possible [KGK+04]. The generic task types in NARA are divided into four sections including task execution, ensuring correct plant status and availability of plant resources, alarm or indication response, and communication.

With the HEP data collected in these three databases, human behavior levels in SRK framework could be quantified when the potential human errors described in these databases are reasonably identified and classified.

3.3.2 Identification and classification of human errors

The identification of human errors is an important phase in HRA. It breaks down the human activities into a more detailed level by task analysis, so that the identification of human errors become possible. It could be either a quantitative or qualitative analysis. The quantitative task analysis requires sufficient data to quantify the probability of errors. The qualitative task analysis could assist in understanding potential human errors.

In [Rea90], the SRK framework combining with human error theory distinguishes human errors into skill-based errors (slips and lapses), rule-based mistakes as well as knowledge-based mistakes. Eight dimensions are discussed to distinguish these three level errors. The distinctions are summarized in Table 3.2 providing suitable references. Operation errors can be classified from the eight dimensions listed in Table 3.2. These eight dimensions of errors contribute to the establishment of the generic error modeling system (GEMS), which is a structured map for detailed examination of the types of errors applicable to the task [Str19].

Furthermore, Hanaman decision tree could be adopted as the joint approach to classify human operation errors into SRK levels [JSXG10]. In the Hanaman decision tree, six influence factors (operation type, crew's understanding of situation,

Table 3.2: Summarized distinctions between skill-based, rule-based and knowledge-based errors (adapted from [Rea90])

Dimension	Skill-based errors	Rule-based errors	Knowledge-based errors
Type of action	Routine actions	Problem-solving activities with rules or knowledge	
Focus of attention	On something other than the task in hand	Directed at problem-related issues	
Control mode	Mainly controlled by automatic processors (Schemata)	(Stored rules)	Limited, conscious processes
Predictability	Largely predictable "strong-but-wrong" errors (Actions)	(Rules)	Variable
Ratio of error to opportunity for error	Though absolute numbers may be high, these constitute a small proportion of the total number of opportunities for error		Absolute numbers small, but opportunity ratio high
Influence of situational factors	Low to moderate; intrinsic factors (frequency of prior use) likely to exert the dominant influence		Extrinsic factors likely to dominate
Ease of detection	Detection usually fairly rapid and effective	Difficult, and often only achieved through external intervention	
Relationship to status change	Knowledge of change not accessed at proper time	When and how anticipated change will occur unknown	Changes not prepared for or anticipated

requirement of procedure, availability of procedure, crew's understanding of procedure, crew's familiarity of procedure) are selected to determine human operation error levels. The structure of Hanaman decision tree is shown as Figure 3.3. From Figure 3.3, it is clear that Hanaman decision tree collects the relationship between influence factors and SRK framework. The meaning of the branches in Hanaman decision tree is explained in Table 3.3. In this case, with Hanaman decision tree, the human error levels could be determined when the states of influence factors are known.

As a short summary, the steps to classify human errors from the three databases into SRK levels are as follows:

1) Understanding the environment of the operation scenario and the characteristics of the operation behavior, identification of the operation tasks including the specific operation steps, working conditions, the time budget for the task, and number of simultaneous actions, etc.

2) Matching the operation behavior characteristics with the listed dimensions in Table 3.2 and the influence factors in Table 3.3

3) Determination of the behavior level of the analyzed task

4) Calculation of the HEP of operation behavior with the consideration of operating environment and working conditions

5) Summarizing the HEP intervals of SRK levels of each database

6) Based on the HEP intervals of SRK levels obtained from the databases, a final HEP intervals table is obtained by calculating the mean HEP of the same level behaviors from the three databases.

The summarized HEP intervals of each task from three databases are listed in Table 3.4, Table 3.5, and Table 3.6. With these three tables, the HEP intervals for skill-based, rule-based, and knowledge-based behaviors can be summarized as Table 3.7. Therefore, the behavior levels in SRK framework can be characterized quantitatively. To visually represent the relationship of HEP between different behavior levels, the principal relations based on numerical values from literature based on the HEP (Table 3.7) is illustrated as shown in Figure 3.4 where x-axis is indicating HEP values and y-axis is presenting cognitive behavior mode (CBM).

3.3.3 Case illustration

For better illustration how human errors are identified and classified into SRK framework and how HEPs of tasks in databases are determined, case illustration from THERP [SG83] is presented.

Table 3.3: The meaning of branches in Hanaman decision tree (adapted from [JSXG10])

Branches	Operation type	Crew's understanding of situation	Requirement of procedure	Availability of procedure	Crew's understanding of procedure	Crew's familiarity of procedure
Upper	Routine	Understanding	Not required	Available	Understanding	Familiar
Lower	Non-routine	Not understanding	Required	Unavailable	Not understanding	Unfamiliar

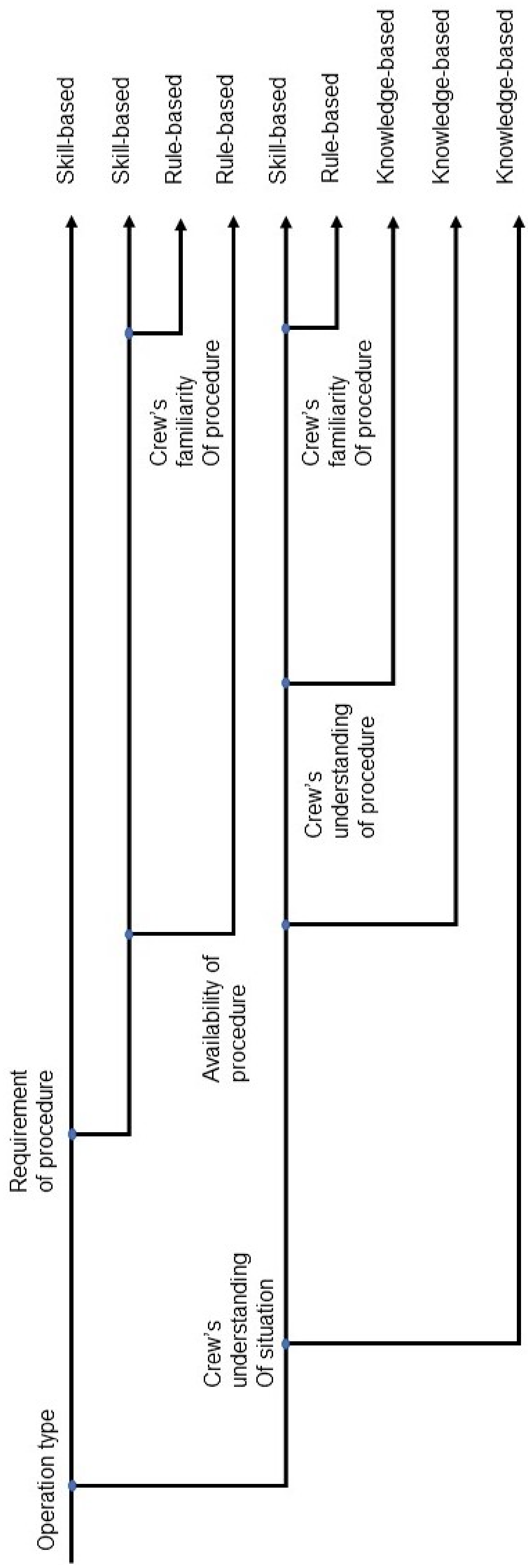

Figure 3.3: Hanaman decision tree (adapted from [JSXG10])

Table 3.4: Summary of skill-based errors and corresponding HEP[He22b]

Databases	Skill-based errors	HEP
THERP	Preparation of written material	3×10^{-3}
	Initiate scheduled shift checking in administrative control	1×10^{-3}
	Using written operation steps in administrative control	5×10^{-3} - 5×10^{-2}
	Omission of item when procedure with checkoff provision use correctly	1×10^{-3} - 3×10^{-3}
	Recalling oral instruction items ont written down	1×10^{-3}
	Selection of unannunciated displays for quantitative or qualitative readings	5×10^{-4} - 3×10^{-3}
	Reading and recording from various numerical indicators	1×10^{-3} - 6×10^{-3}
	Check-reading from various types of displays	1×10^{-3} - 6×10^{-3}
	Inadvertent activation of a control	3×10^{-2}
	Turn a rotary control or two-position switch with common stereotype in wrong direction	1×10^{-4} - 5×10^{-4}
	Set rotary control to an incorrect setting	1×10^{-3}
	Fail to complete change of state of a component which must be held until change is complete	3×10^{-3}
	Improperly mate a connector	3×10^{-3}
	Selection in changing or restoring a locally operated valve from a group of unambiguously labeled valves which are set apart from similar looking valves	1×10^{-3} - 3×10^{-3}
	Detect stuck locally operated valves with indications are available	1×10^{-3} - 5×10^{-3}
	Resume attention to a legend light within 1 minute after an interruption	1×10^{-3}
SRS-HRA	Communication error	5×10^{-2}
	Incorrect labeling or tagging	5×10^{-3}
	Failure to lock out	5×10^{-4}
	Chemical addition or elution error	3×10^{-3}
	Transfer error	$3 \times 10^{-6}/tank - h$
	Overfilling of a tank	$5 \times 10^{-6}/tank - h$
	Laboratory analysis error	3×10^{-4}
	Random actuation/shutdown of system	$5 \times 10^{-6}/h$
	Vehicle collision with stationary object	$1 \times 10^{-6}/miles$
	Single vehicle accident	$1 \times 10^{-6}/miles$
	Vehicle collision with another moving vehicle	$1 \times 10^{-6}/miles$
	Dropping of load when using forklift	5×10^{-5}
	Puncturing of load when using forklift	3×10^{-5}
	Dropping of load when using crane/hoist	1×10^{-4}
	Crane/hoist strikes stationary object	3×10^{-4}
NARA	Carry out simple single manual action with feedback	5×10^{-3}
	Perform completely familiar, well designed, highly practiced, routine task	1×10^{-4}
	Set system status as part of operations using strict administratively controlled procedures	7×10^{-4}
	Calibrate plant equipment using procedure	3×10^{-3}
	Simple response to a range of alarms or indication providing clear indication of situation	4×10^{-4}
	Verbal communication of safety-critical data	6×10^{-3}

Table 3.5: Summary of rule-based errors and corresponding HEP [He22b]

Databases	Rule-based errors	HEP
THERP	Rule-based actions by control room personnel after diagnosis of an abnormal event	2.5×10^{-2} - 5×10^{-2}
	Carry out a plant policy or scheduled task	1×10^{-2}
	Use a valve change or restoration list in administrative control	1×10^{-2}
	Use a written test or calibration procedure in administrative control	5×10^{-2}
	Omission per item when procedure without provision are used or incorrectly used	3×10^{-3} - 1×10^{-2}
	Arithmetic calculation errors	1×10^{-2} - 5×10^{-2}
	Selection of control on a panel from an array of similar-appearing controls	5×10^{-4} - 3×10^{-3}
	Turn rotary control or two-position switch with unusual stereotype in wrong direction	1×10^{-2} - 5×10^{-1}
	Select wrong circuit breaker in a group of circuit breaker	3×10^{-3} - 5×10^{-3}
	Selection in changing or restoring a locally operated valve from group of ambiguously labeled and similar appearance of valves	5×10^{-3} - 1×10^{-2}
	Checker checks non-routine task or involve active participation	1×10^{-2} - 5×10^{-2}
	Checking the status of equipment if that status affects one's safety either by checker or maintainer	5×10^{-4} - 1×10^{-3}
	Response to multiple annunciators alarming closely in time	1×10^{-4} - 5×10^{-2}
SRS-HRA	Failure of administrative control	5×10^{-3}
	Failure to verify within control room	1×10^{-2}
	Failure to verify outside control room	3×10^{-2}
	Error in selecting control within control room	1×10^{-2}
	Error in selecting control outside control room	1×10^{-2}
	Incorrect reading or recording of data	1×10^{-2}
	Miscalibration	5×10^{-3}
	Failure to restore following test	1×10^{-2}
	Failure to restore following maintenance	5×10^{-3}
	Failure to verify parameter with calculation	3×10^{-2}
	Excavation error	1×10^{-2}
	Failure of long-term accident recovery	3×10^{-3}
NARA	Start or reconfigure a system from the main control room following procedures, with feedback	1×10^{-3}
	Start or reconfigure a system from a local control panel following procedures, with feedback	2×10^{-3}
	Routine check of plant status	2×10^{-2}
	Restore a single train of system to correct operational status after a test, following procedures	4×10^{-3}

Table 3.6: Summary of knowledge-based errors and corresponding HEP [He22b]

	Knowledge-based errors	HEP
THERP	Diagnosis of the abnormal events within certain time	5×10^{-1}
	Rule-based actions by control room personnel after diagnosis of an abnormal event	1.0
	Perform the task without using written maintenance procedures or checklist	3×10^{-1} - 5×10^{-1}
	Written procedures are available and should be used but not used	5×10^{-1}
	Reading and recording from various large number of parameters recorder and graphs	1×10^{-2} - 5×10^{-2}
	Recognize that an instrument being read is jammed without indicators to alert the user	1×10^{-1}
	Detect stuck locally operated valves when indications are not available	1×10^{-2}
	Checker checks routine tasks with or without written materials	1×10^{-1} - 2×10^{-1}
	Checker notices the locally operated valve is not completely opened or closed after the valve is checked	1×10^{-1} - 9×10^{-1}
	Checking the task in a two-man team	5×10^{-1}
	Respond to a legend light if more than 1 minute elapses after an interruption	9.5×10^{-1}
	Respond to a steady-on legend light at initial audit or hourly scans	9×10^{-1} - 9.5×10^{-1}
	Fail to detect unannunciated deviant display	9.5×10^{-1} - 9.9×10^{-1}
	Fail to detect multiple unannunciated deviant displays	1×10^{-3} - 9.9×10^{-1}
	Daily walk-around inspection	5.2×10^{-1}
SRS-HRA	Failure to respond to compelling signal	1×10^{-2}
	Checker verification error	1×10^{-1}
	Supervisor verification error	3×10^{-1}
	Diagnosis error	1×10^{-2}
	Failure of visual inspection	1×10^{-1}
	Failure of manual fire detection	1×10^{-1}
	Failure of manual fire suppression by occupant	3×10^{-1}
	Failure of manual fire suppression by non-occupant	3×10^{-1}
	Failure of long-term accident recovery	3×10^{-3}
NARA	Judgment needed for appropriate procedure to be followed based on interpretation of a situation	6×10^{-3}
	Carry out analysis	3×10^{-2}
	Identification of situation requiring interpretation of complex patter of alarms or indications	2×10^{-1}

Table 3.7: HEP intervals for three level errors [He22b]

Databases	Skill-based error	Rule-based error	Knowledge-based error
THERP	1×10^{-4} - 5×10^{-3}	1×10^{-4} - 5×10^{-2}	1×10^{-3} - 1.0
SRS-HRA	3×10^{-5} - 5×10^{-3}	3×10^{-3} - 3×10^{-2}	3×10^{-3} - 3×10^{-1}
NARA	1×10^{-4} - 6×10^{-3}	1×10^{-3} - 2×10^{-2}	6×10^{-3} - 2×10^{-1}
Mean	7×10^{-5} - 5.3×10^{-3}	1×10^{-3} - 3.3×10^{-2}	3×10^{-3} - 5×10^{-1}

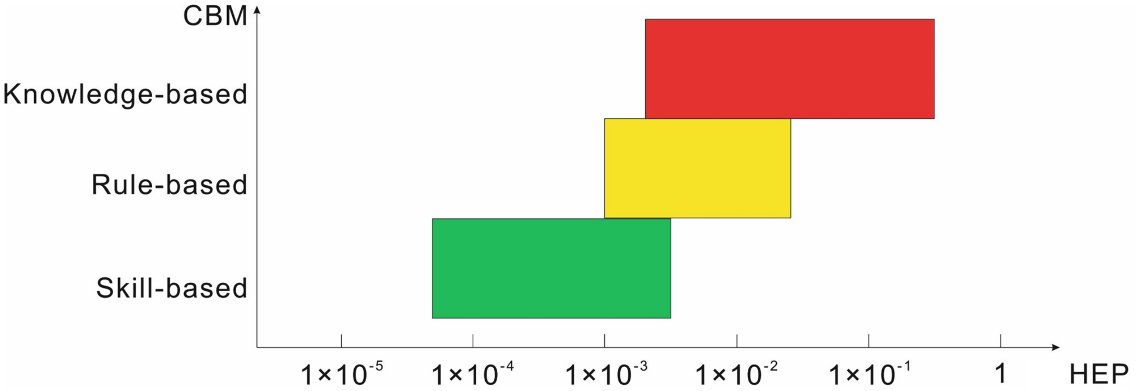

Figure 3.4: Relationship between human behavior levels and HEP [He22b]

For demonstration, the selected task is chosen as administrative plant control. It refers to the extent the plant is run in conformance to the guidelines by which it was designed to operate, reflects the type of structure inherent in a plant, and reinforces the lines of responsibility. The human operators involved are responsible for the performance of certain tasks necessary to reliable and safe plant operation in both normal and abnormal situations. The possible failures in administrative control with the estimated HEPs are listed in Table 3.8.

For item (1), plant policy refers to a set of operating requirements that plant management generally expects to be followed. These structured requirements are described in a formal set of written instructions that are available to all operation staffs in relevant positions. The estimated HEP is assigned to be 0.01 (EF = 5). In this case, the upper bound of HEP of task item (1) is 0.05, while the lower bound of which is 0.002. Therefore, the HEP of task item (1) can vary within the range between 0.002 to 0.05. With the description of tasks item (1), the error level of this item could be classified by the joint approaches of eight dimensions in Table 3.2, and the information from the Hanaman decision tree. The determination of error level with eight dimensions for item (1) and its explanation is presented in Table 3.9. Next the Hanaman decision tree is applied for the double check. The operation type is 'routine', so it is the upper branch; the requirement of procedure is 'required', so it goes to the lower branch; the availability of procedure is 'unavailable' because operators did not follow the procedures, so it is the lower branch. In this case, the item (1) is 'rule-based error' from the Hanaman decision tree. So the HEP of item (1) belongs to 'rule-based level'. In this study, the effect of EF on HEP is not considered as EF represents the upper and lower bound of HEP for special cases, and the nominal HEP is for most of the common cases.

The results from the eight dimensions approach and Hanaman are almost identical but has to be adapted in detail in some cases when the results are different. In this case the results from eight dimensions will be mainly adopted because this approach has more degrees of freedom to be adapted.

Table 3.8: Estimated HEPs related to failure of administrative control (adapted from [SG83])

Item	Task	HEP	EF
(1)	Carry out a plant policy or scheduled tasks such as periodic tests or maintenance performed weekly, monthly, or at longer intervals	1×10^{-2}	5
(2)	Initiate a scheduled shiftly checking or inspection function	1×10^{-3}	3
	Use written operation procedures under		
(3)	Normal operating conditions	1×10^{-2}	3
(4)	Abnormal operating conditions	5×10^{-3}	10
(5)	Use a valve change or restoration list	1×10^{-2}	3
(6)	Use written test or calibration procedures	5×10^{-2}	5
(7)	Use written maintenance procedures	3×10^{-1}	5
(8)	Use a checklist properly	5×10^{-1}	5

As a final result, the new introduced Figure 3.4 illustrates the relationship between cognitive behavior modes and HEP values. For the first time here the SRK levels are mapped with HEP values. This builds now the base for consideration of additional effects related to time pressure and training levels which is discussed in chapter 3.4.

3.4 Analysis and application

In [Vic99], the effect of switching behavioral levels was investigated. Knowledge-based activities can be executed 'online' and synchronously, which means that the whole process is realized 'online' using skills or rules (or as a tool). For example, a pilot manually controls an aircraft using skill-based behaviors while simultaneously applying knowledge-based behaviors to decide whether the target inclination is appropriate [FP16].

It should be recognized that the switching between SRK levels can be identified as short or long time scaled. Switching between SRK levels can be realized in short time (some activities mainly refer to the 'online' activities which require real-time feedback). The time scale could be seconds, minutes, or hours, depending on the

Table 3.9: Determination of error level with eight dimensions for item (1) and its explanation [He22b]

Dimension	Item (1)	Explanation
Type of action	Routine actions	Plant policy is described fully in a formal set of written instructions that operators need to follow in the operation procedures.
Focus of attention	On something other than the task in hand	The related errors in item (1) as described include the operator's attention may be drawn away from the tasks at hand by another, more compelling demand for action.
Control mode	Mainly by automatic processors (stored rules)	Operators need to follow the rules for operation. They are not always check the steps with the checklist, most of the steps are memorized.
Predictability of error types	Largely predictable "strong-but-wrong" errors (rules)	The errors of plant policy or scheduled tasks are predictable as the operations are followed by controlled rules and the steps are clear.
Ratio of error to opportunity for error	Absolute numbers may be high, but constitute a small proportion of total number of opportunities for errors	The absolute numbers may be high because they are daily, weekly operations, when considering the error rate, it is low as the operation frequency is high in daily, weekly working.
Influence of situational factors	Low to moderate; intrinsic factors likely to exert the dominant influence	Not following the plant policy or did not have the periodic tests may have low effect on the safety production of plant, but when the error at some part is not detected because lack of periodic tests, which leads to large failures or accidents, it will exert the dominant influence.
Ease of detection	Detection usually fairly rapid and effective	The errors can be detected when checking the taggings and checklists, or discuss with the responsible operators to know his familiarity on the duties.
Relationship to status change	When and how anticipated change will occur unknown	Due to the lack of periodic tests, it is known that status change will happen, but when and how it will happen is unknown.
	Rule-based errors	

situations. The skill-based behaviors related to highly routine activities in familiar environment. Rule-based behaviors are involved when attention checks upon progress and detects a deviation from the planned-for conditions. When operators realize that their rule-based solutions are not able to cope with the problem, knowledge-based performance is engaged. The activities of knowledge-based levels can be stopped when adequate plans for problem are acquired, which leads to the rule-/skill-based behaviors again.

When the time is stretched to weeks, months or years, the SRK level of operator behaviors could be switched depending on their experience regarding the situations they encountered. Operators who have continuous training on specific situations will increase their experience, which lead to behavior level switch from knowledge-based to skill-/rule-based. Meanwhile, after a long period of no training, the experience that operators previously occupied may be lost, thereby changing their behaviors from skill-/rule-based to knowledge-based.

3.4.1 Effects of time pressure and training on SRK levels switching

Two performance shaping factors (PSFs) namely time pressure and training are selected for the analysis of SRK level switching as these two PSFs affect SRK level switching and human reliability of operators significantly from short time scale and long time scale, respectively.

Time pressure

Time pressure has strong effects on human judgment and decision making as the strategies of coping with situations under time pressure are changed comparing with non-time pressure situations [ME97]. For example, the strategy of acceleration may be adopted with a faster rate of information processing and/or reducing pause and other interruptions in task-related activities. Filtering (processing some parts of the information more, and others less), acceleration, and omission (ignore particular parts of the information) are mostly employed strategies by human operators to deal with time pressure situations [SM93]. The relationship between time pressure and human performance is in inverted U-shape (increasing time pressure could induce to better human performance up to a certain point). After this point, human performance is decreasing with more time pressure [Hwa94]. In different application fields, time pressure is a key factor affecting human reliability, which often cause premature decision making, increased risk tolerance, and impaired cognitive performance and health [HMB+09]. In transportation, time pressure is regarded as the most hazardous task characteristics of emergency vehicle driving [HCS18]. In aviation maintenance, time pressure is the most frequently mentioned factor leading to incidents from a survey as maintenance operators tempt to take shortcuts to get an aircraft back into service more quickly [RH17]. Air traffic control (ATC) is characterized by time pressure, multiple tasks and goals, and high error consequences because continuous increasing in the volume of air traffic imposes more demands on air traffic controller [KM13].

Training

Training helps to enhance human operator performance, so to reduce human errors. Whenever a human operator's ability to perform a task is limited by lack of knowledge or skill, it is making sense to bridge the gap by training [SB15]. Training is one of the essential constituents of a quality system process, delivering qualified operators to meet the demands of exacting roles. Ineffective or negative training is a major latent failure in the overall accident causation chain [SMC10]. From [BC10], it can be identified that sub-optimal training is one of the two most critical flight hazards in aviation, with the other one being a shortage of experienced operators. Training of control room operating crews in nuclear power plant consists of two stages: one is a lengthy process of initial training in which acquiring knowledge on appropriately carrying out the tasks to be performed in the control room, the other is a continuous training aimed maintaining and improving the knowledge and skills on operation [Dia11]. Training standards need to be established for partially automated vehicles as driver assistance systems (ADAS) become standard equipment for lower-priced vehicles [CH19]. Hence, the effects of training on operator performance should be monitored and measured to identify the effectiveness on human reliability enhancement.

It is mentioned that time pressure and training as two PSFs deeply affect human reliability. When SRK framework is considered in the effects analysis of time pressure and training on SRK levels switching, the map (Figure 3.4) generated in this work could be applied to visualize the effects. In Figure 3.5 (a), the effects of time pressure on SRK levels switching are indicated. The detailed explanation of the switching behaviors is presented in Table 3.10. The effects of training on SRK level switching is presented in Figure 3.5 (b). The detailed explanation of switching behaviors is shown in Table 3.11. It is obtained from Figure 3.5 that human performance could be switched not only among the same level, but also between levels with different extent of training.

3.4.2 Framework of the SRK levels switching

Although autonomous and semi-autonomous systems are applied to different application fields, human operators are still the center for human-machine systems regarding safety issues. Even the most advanced automated systems still need humans to monitor the situations and takeover or stop the system when emergencies

Table 3.10: Explanation of SRK levels switching with time pressure [He22b]

Item	Explanation	References
[P1], [P2]	In routine tasks and very familiar environment, HEP of operator may decrease if time pressure is at less/appropriate degree ([P1]), or it may increase if time pressure is at appropriate/greater degree ([P2]).	[Hwa94], [CSM15]
[P4], [P5]	In rule-based tasks, human operator may take the strategies of acceleration or omission with highly time pressure, which will increase HEP ([P5]); when less time pressure imposed, operator has more time to follow the established rules, so to reduce HEP ([P4]).	[Hwa94], [HLL92]
[P8], [P9]	In knowledge-based tasks, human operator may not able to search for additional solutions for the problem as there is not enough time ([P9]); when operator is with less time pressure, some solutions may be found ([P8]).	[Hwa94], [HLL92]
[P3], [P6]	With appropriate time pressure not exceeding the inverted U-shaped top point, human behavior could switch between skill-based level and rule-based level. For routine tasks in familiar situations, when appropriate time pressure is imposed, rule-based behaviors could move to skill-based as human performance is increasing ([P3]); while skill-based behaviors will move to rule-based with little time pressure ([P6])	[Hwa94], [CSM15]
[P7], [P10]	With appropriate time pressure not exceeding the inverted U-shaped top point, human behavior could switch between rule-based level and knowledge-based level. When appropriate time pressure is imposed, knowledge-based behaviors could move to rule-based because of increased performance ([P7]). Rule-based behaviors will move to knowledge-based with less time pressure ([P10])	[Hwa94], [AG03]

Table 3.11: Explanation of SRK levels switching with training [He22b]

Item	Explanation	References
[T1], [T2]	Human operator may be more skilled on routine tasks with continuous training ([T1]), or the familiarity on skilled tasks decrease because of short time leaving the tasks ([T2]).	[DC96]
[T4], [T5]	For rule-based tasks, human performance may be improved with more training ([T4]), and may be reduced with less training ([T5]).	[CGW13]
[T8], [T9]	For unknown situations with no know-how or rules for tasks, human reliability is increased with training on knowledge-based tasks ([T9]), and is decreased with less training ([T8]).	[CGW13]
[T3], [T6]	Human performance may be switched from skill-based to rule-based when lack of training ([T6]); with more training, rule-based behaviors could be moved to skill-based ([T3]).	[MPN+11]
[T7], [T10]	With more knowledge and training with operation situations, human performance could be moved to rule-based level from knowledge-based ([T7]); when leaving the tasks for long time, human's experience may be lost, so the performance may be switched from rule-based to knowledge-based ([T10]).	[MPN+11], [CGW13]

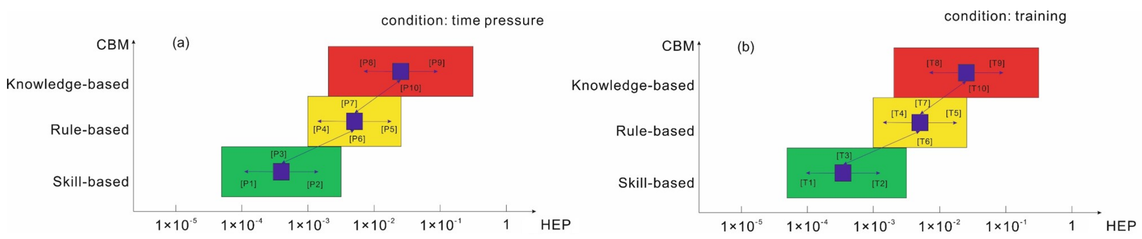

Figure 3.5: Effects of time pressure (a) and training (b) on levels switching (in combination with the numerical values for Ti and Pi) [He22b]

occur. At the same time, high automation may increase boredom and decrease vigilance which affects the ability to takeover control of the system [MSJ$^+$15]. Therefore, quality of human performance is critical to the reliability of human-machine systems. Many measures have been developed to monitor human performance in human-automation systems, especially human-driving system. In [DHUM10], five types (subjective report measures, driver biological measures, driver physical measures, driving performance measures, and hybrid measures) of driver inattention monitoring measures are summarized. In [FHLR20], human factors regarding automated vehicles, such as the workload, distraction, situation awareness (SA) and driver trust, are discussed. The ultimate question to be answered by these studies of human factors is the monitoring and evaluation of human reliability. The quantitative study of human behavior reliability of different levels in SRK framework discussed in this thesis provides possibilities for the evaluation of human errors and human reliability. Meanwhile, the study of the effects of time pressure and training on the levels switching demonstrates the dynamic changes of SRK framework for environment. Hence, a more general structure to illustrate the dynamic behavior of levels switching could be established which is illustrated in Figure 3.6.

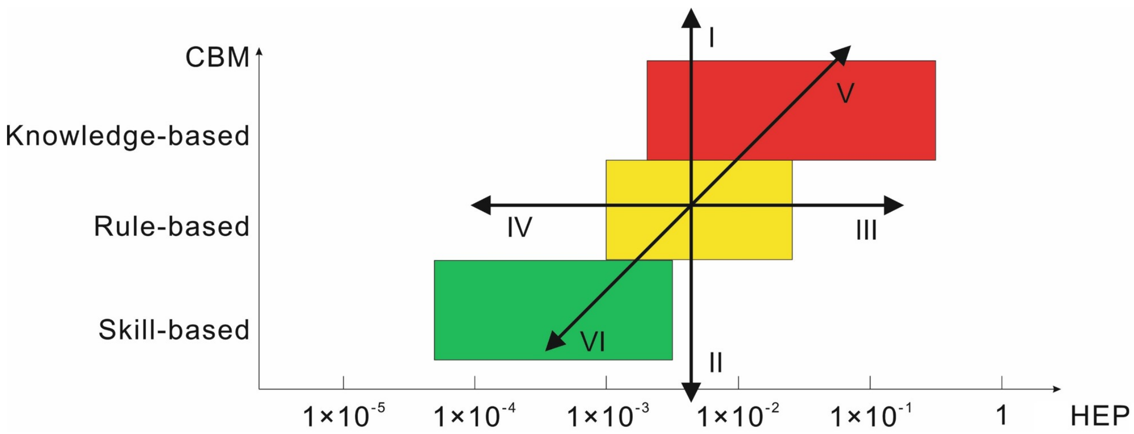

Figure 3.6: Analysis of the dynamic behavior of SRK levels switching [He22b]

From Figure 3.6, it can be obtained that six directions are used to indicate the relationship between HEP and human behaviors.

Directions I/II mean that the quality of the tasks is different but HEP is identical. The typical example for this case is that a very experienced operator is in process of tasks in familiar environment, although some rules are available or situations need to be diagnosed and new plan should be generated, human reliability is identical as the solutions could be found easily.

Directions III/IV indicate that the quality of tasks is identical, while human operators' experience level is varying. As the experience levels of human operators on situations are different, their human performance ability regarding the same task is also different, which induces the HEP varying.

Directions V/VI present loosing experience (V) and typical learning process (VI). As learning continues, operators become more familiar with the situations and more proficient in the process, so the behavior level switches from knowledge-based level and eventually to skill-based level. Meanwhile, after long time of decoupling from the operation loop (due to automation) or specific tasks (due to tasks changing), human operators will loose abilities for the tasks, so their experience on tasks is gradually fading away (knowledge degradation) and the behaviors finally reach knowledge-based level.

Many approaches and techniques have been developed for human performance assistance to reduce risks in application fields. In [NNAV21], the technologies for driver assistance system (ADAS) driven solutions are summarized, the eye-gaze and head pose estimation in vision intelligence are reviewed and the development of learning algorithms makes it possible to develop a real-time recommendation system for autonomous vehicle. In aviation, the human performance model is used to improve predictions of situation awareness of pilots [HGW+11]. In [FZBD+18], the impact of seafarers' emotion on their performance is investigated with electroencephalogram (EEG) and self rating. In the previous work of the authors, human driver reliability is evaluated using a modified fuzzy-based CREAM approach with the data collected from driving simulator [HLLS21]. The approach of estimation of human reliability could be developed into a real-time monitoring system for human driver. When the driver displays low human reliability in some situations, the system could issue alerts to bring the driver's attention and ability back to the driving operation. In some critical situation, when human driver reliability is extremely low, and the vehicle cannot be controlled at all, the system could directly takeover the vehicle from the driver. Hence, the work of quantification of human behavior levels regarding SRK model lays the foundation of evaluation between automation and operator's takeover.

The framework of SRK levels switching considering the effects of time pressure and training provides the idea for evaluation of simulator training for daily tasks for individuals. When simulator training data are collected, human operators' reliability could be estimated based on the approach proposed in [HLLS21], which could be quantified into points mapping into Figure 3.6. In this case, the actual training status of individuals could be recognized and training suggestions for further steps could be made. Meanwhile, the error types could be also identified by the map, which helps to analyze and improve human performance during training.

3.5 Summary

Humans are always somewhere integrated in the loops although the automation level in human-machine systems is getting higher with the development of technology. Human error is causing an increasing proportion of total accidents. With the research on human error mechanisms and failure modes, the study of human reliability analysis (HRA) has been formed. Many cognitive process models have been established to explain human performance. Among these models, skill-, rule-, knoledge-based behavior (SRK) model is wildly used. In this chapter, human behavior levels of SRK framework are quantified and the effects of two performance shaping factors, time pressure and training, on levels switching are analyzed. Based on the analysis a new graphical summary is developed to illustrate the effects. The main work could be summarized as follows:

1. The HRA methods, including the so called 'first generation', 'second generation' and 'third generation', are briefly discussed, the cognition process in these HRA methods are summarized, and SRK model is selected in this thesis to characterize human behavior.

2. Three level behaviors in SRK model are illustrated. With the description of three levels, it could be concluded that skill-based behaviors relate to the higher human performance reliability and knowledge-based behaviors correspond to the lowest, but the defining of HEP values of each level need human reliability data research.

3. Human error probability (HEP) from three databases (THERP, SRS-HRA and NARA) are collected to quantify human behavior levels in SRK model. The detailed procedures for the identification and classification of human errors are illustrated. A case study regarding classifying the task of administrative control in plants into SRK levels is presented to explain how the procedure works. Finally, the HEP intervals of SRK levels are summarized and a graphical framework presenting the relationship between human behaviors and HEP is generated.

4. The effects of time pressure and training on SRK levels switching are analyzed and the switching behaviors are explained. Human behavior levels in SRK model can switch in several ways. The switching behaviors could be identified as short time scale and long time scale. Short time scale switching mainly

refers to the 'online' activities where real-time feedback is required. Human behavior levels of the SRK framework can be switched depending on experience regarding the tasks and environment for the long time scale. Two performance shaping factors including time pressure and training are selected for analysis of SRK levels switching. It is obtained that human behaviors can be switched between levels with time pressure and training. It can be stated that the established visual connections show the effects with respect to time pressure and additional training. Furthermore, it becomes clear that from the HPE point of view, the SRK levels roughly correlate but in detail overlap.

5. A general map describing SRK levels switching with six different directions is generated, the explanation of each direction is presented. The new graphical illustration allows: i) a human performance reliability monitoring system to be established combining with the fuzzy-based modified CREAM approach in this thesis; ii) the individual recognition and evaluation system of training status to be generated with collected operator training data.

4 Human reliability estimation in dynamic context

In this chapter, the methods adopted in modified fuzzy-based CREAM approach approach including CREAM approach, fuzzy theory, data clustering approach (FN-DBSCAN, CLUSTERDB, GMFPE algorithms) are explained and applied to driving data for data clustering. The human driver evaluation concept HPRS (Human performance reliability score) is introduced. To depict the dynamic driving context, the new list of CPCs is established.

Part of the contents, figures, and tables presented in this chapter are modified after previous publications [HS22][HBS22][HLLS21][HLL21][HS20][HTS20][He19]. Part of the contents, figures, and tables are prepared for publication of [He22a].

4.1 CREAM approach

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 [ZWX+17].

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) [CRSWP96], and cognitive reliability and error analysis method (CREAM) [Hol98].

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.

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 [Hol98]:

- 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 Table4.1. The reliability intervals (probability of action failure) of control modes come from statistical data in industries.

Table 4.1: CPC control modes and their probability interval (adapted from [Hol98])

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)

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, 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 Figure 4.1. 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.

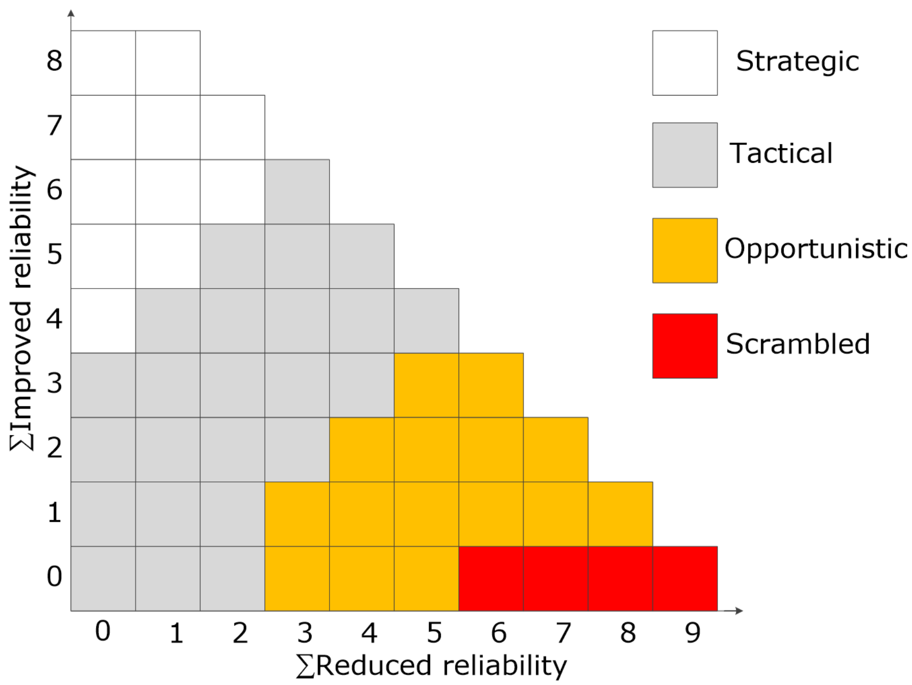

Figure 4.1: Relations between CPC score and control modes (adapted from [Hol98])

4.2 Fuzzy theory

Fuzzy logic is to model the imprecise modes of reasoning that are fundamental for the capacity of rational decision-making in an environment of uncertainty and imprecision [Háj13]. Fuzzy logic is based on the degree of truth of a logically compound proposition which can obtain the value between 0 an 1 rather than the value of truth (1) or false (0) in standard Boolean logic. Fuzzy logic is established by Lotfi Zadeh based on his earlier work on fuzzy set theory [Zad78]. In [Zad78], a fuzzy set A in the universe of discourse X, is defined by a membership function μ_A, which correlates each element x in X to a real number in the interval [0,1], where the degree of membership of x in A is denoted by the value of μ_A.

The main features of a membership function are the height, core, and support parameters. The height of a fuzzy set A can be represented with the mathematical function

$$height\,(A) = max\{\mu_A(x)|x \in X\}, \tag{4.1}$$

which indicates the highest value of the membership function. The domain of height can be any value in the range of 0 to 1. The core of the membership function can be defined mathematically by

$$core\,(A) = max\{x|x \in X, \mu_A{=}1\}, \tag{4.2}$$

where the core contains all elements x which are characterized by full membership in the set, in this case with a value of 1. The support of a membership function of a fuzzy set A can be expressed by

$$supp\,(A) = max\{x|x \in X, \mu_A(x) > 0\}, \tag{4.3}$$

where the support contains all elements x which are characterized by a nonzero membership in the set.

Trapezoidal membership function is selected to describe the membership degree of CPCs. The advantages of its simplicity and popularity make the trapezoidal membership function mainly used [DHO08].

4.3 Data clustering approaches

Clustering is a method to group similar objects in the same group and separate dissimilar objects into different groups [SSBD14]. With fuzzy theory, data clustering approaches are used to find the core points and support points of the data. In this case, the membership functions could be defined by data rather than by expert knowledge. In this thesis, three data clustering approaches are adopted to define the membership functions.

4.3.1 FN-DBSCAN

In the standard DBSCAN approach, classical neighborhood density analysis is applied to determine the core points of clusters. A core point is defined if the number of points in a specific radius is larger than a certain threshold [NU09]. The FN-DBSCAN algorithm implements fuzzy neighborhood cardinality to generate core points [UN08]. To define the membership functions, the cores and supports of a trapezoidal membership function are determined using the core and support points based on the fuzzy density-neighborhood of the centroid of clusters.

The fuzzy neighborhood membership function could be defined as

$$N_x(y) = \begin{cases} 1 - \frac{d(x,y)}{d^{max}} & \text{if} \quad d(x,y) \leq \epsilon, \\ 0 & otherwise, \end{cases} \tag{4.4}$$

where $d(x, y)$ represents the distance between any points x and y, whereas ϵ determines the maximal threshold of the distance between points.

To further improve the sensitivity of the points with different distances to the neighbor points, the neighborhood membership functions dependent on the parameter k is expressed as

$$N_x(y) = max\{1 - k\frac{d(x,y)}{d^{max}}\}. \tag{4.5}$$

The fuzzy neighborhood set of point $x \in X$ with parameters ϵ_1 is expressed as

$$FN(x; \epsilon_1) = \{< y, N_x(y) > \mid y \in X, N_x(y) \geq \epsilon_1\}, \tag{4.6}$$

where ϵ_1 defines the minimal threshold of the neighborhood membership degree, N_x refers to any membership function that describes the neighborhood relation between points.

A point x is defined as a fuzzy core point with parameters ϵ_1 and ϵ_2 if it fulfills the requirement of

$$cardFN(x; \epsilon_1, \epsilon_2) \equiv \sum_{y \in N(x;\epsilon_1)} N_x(y) \geq \epsilon_2, \tag{4.7}$$

Nonetheless, the parameters ϵ_1, ϵ_2, and k must still be defined using expert knowledge. Further details regarding ϵ_1 and ϵ_2 are given in [UN08]. To decrease the dependency of parameters on expert knowledge, it is suggested a way to reduce the pre-defined parameters from 3 to 1. ϵ is defined as the average distance between adjacent data

$$\epsilon = \frac{\sum_{i=1}^{m-1} d(x_i, x_{i+1})}{m-1}, \tag{4.8}$$

where $d(x_i, x_{i+1})$ represents the distance between the i-th data point and its adjacent neighboring data point while m is the total number of data points. Therefore, if a data point y is closer to a data point x than the average distance between adjacent data, data point y then has a neighborhood degree of $N_x(y) > 0$.

The relation of parameter k with respect to ϵ can be represented by

$$k = \frac{d^{max}}{\epsilon}. \tag{4.9}$$

Furthermore, parameter ϵ_1 defines the radius of the membership threshold of data points to be included in the fuzzy cardinality is given the value of 0. As a result of ϵ in Eq. 4.8, a neighborhood consisting of relatively close points is considered and therefore these data points could also be included in the fuzzy cardinality. Given that $\epsilon_1 > 0$, the density requirement towards the center of the neighborhood could be increased. Thus, only fuzzy cardinality threshold ϵ_2 needs to be determined.

To get the optimal value of ϵ_2, the genetic algorithm as explained in [Gre86] is applied. To implement the genetic algorithm to find the optimal ϵ_2 value, the operations of roulette wheel selection, mutation and crossover are used. The chromosomes are comprised of the ϵ_2 values for each variable of the dataset. In order to define the fitness of a chromosome, the data set is fuzzified by the membership functions resulting from the chromosome. The fuzzified data is used to train and test the classification algorithm of KNN (K-nearest neighbors). The classification accuracy of the test data is used as fitness value of the chromosome.

4.3.2 CLUSTERDB*

The CLUSTERDB* algorithm is a density-based approach which extends the CLUSTER algorithm by the validity index DB^* to avoid the generation of small and not well separated clusters [HO07].

The DB^* describes the overall similarity of all cluster nc and is defined as

$$DB^*(nc) = \frac{1}{nc}\sum_{i=1}^{nc}\left\{\frac{max_{k=1,...nc,k\neq i}(S_i + S_k)}{min_{p=1,...nc,p\neq i}d_{ip}})\right\}, \tag{4.10}$$

where d_{ip} is the distance between the i-th and p-th centroid. S_i is the scatter distance

$$S_i = \frac{1}{|C_i|}\sum_{x\in C_i}\|x - c_i\|, \tag{4.11}$$

of the i-th cluster C_i with the centroid c_i.

A similarity value of s needs to be predefined for the number of membership functions. The data is first sorted in ascending order and the distance $diff$ between adjacent data points are calculated. The similarity values between adjacent data can be determined as

$$s_i = \begin{cases} 1 - \frac{diff_i}{N \cdot \sigma_s} & \text{if} \quad diff_i \leq N \cdot \sigma_s, \\ 0 & \quad otherwise, \end{cases} \tag{4.12}$$

where s_i denotes the similarity and $diff_i$ is the distance between data points of x_i and x_{i+1}. The predefined parameter N decides the shape of membership functions and σ_s denotes the standard deviation of all $diff$. With the resulting similarity values, the data set can be partitioned by the use of the predefined similarity threshold α. Adjacent data with similarity values larger than α are considered as among the same cluster.

To automatically generate membership functions, the value of the parameter (α) needs to be optimally defined. The parameter of N does not have to be optimally defined. A change in N changes the similarity value s_i with the same magnitude. In this case, the same genetic algorithm and fitness evaluation method used in FN-DBSCAN is applied to define the optimal value of α and to calculate the fitness.

4.3.3 GMFPE

The GMFPE algorithm is a genetic-based algorithm estimating the optimal trapezoidal membership function parameters with a predefined number of membership functions. The chromosomes contain the trapezoidal core parameters x_l and x_r and the support parameters x_a and x_e of all membership functions. In this case, the total size of chromosomes of all variables is as

$$size = \sum_{i=1}^{k} 4 \times m_i, \tag{4.13}$$

where k is the total number of variables, m_i is the number of membership functions of variable i.

A chromosome is comprised of multiple sections containing the membership function parameters in each variable. The genes within each section are organized so that every first and fourth gene represents x_a and x_e and every second and third gene represents x_l and x_r of a membership function. The genes of each section in a chromosome are generated randomly to generate an initial population. The ranges of gens are between the minimum and maximum of each variable value in data set.

Two-point crossover with two chosen points indicating sections of the chromosome is adopted in GMFPE algorithm [Shi19]. The boundary of mutation of a gene is depending on the parameter type of the gene. The mutations x'_{aj}, x'_{ej}, x'_{lj}, and x'_{rj} of the j-th membership function parameters x_{aj}, x_{ej}, x_{lj}, and x_{rj} are defined as

$$x_{r(j-1)} \leq x'_{aj} \leq x_{e(j-1)}, \tag{4.14}$$

$$x_{e(j-1)} \leq x'_{lj} \leq x_{rj}, \tag{4.15}$$

$$x_{lj} \leq x'_{rj} \leq x_{a(j+1)}, \; and \tag{4.16}$$

$$x_{a(j+1)} \leq x'_{ej} \leq x_{l(j+1)}. \tag{4.17}$$

To evaluate the chromosomes fitness, the same genetic method as described in FN-DBSCAN is applied. The new generation is formed when children with fitness values higher than the lowest fitness values in the current population are exchanged with the respective chromosomes [McC05].

4.4 Human performance reliability score (HPRS)

The traditional HRA methods capture human performance at a particular point in time which could be considered as static HRA methods, so the changes of performance shaping factors (PSFs) affecting each other and the event progression are not discussed [Bor07]. When it comes to the dynamic HRA methods, not only the continuous time-sliced human error probability (HEP) calculation is afforded, but also the modifications of PSFs are available. Table 4.2 presents the modification of PSFs from static conditions to dynamic progression, which indicates that static HRA methods could be modified into dynamic HRA methods when the PSFs evolution with time are defined. Actually, in [Bor07] it is stated that the utilization of cognitive modeling and simulation to produce a framework of data to quantify the likelihood of human error rather than the development of specific HRA methods is the key to dynamic HRA.

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 [HLLS21]. The driving context is dynamically changing in real time. Human driver's ability to perform correctly in the different context requiring different experiences etc. is also fluctuating in real time. In this case, human driver reliability should be evaluated dynamically. When dynamic changing context and the event progression are taken into account, adaptions

Table 4.2: Modification of PSFs (adapted from [Bor07])

Static condition	Dynamic progression	Dynamic initiator
PSFs remain constant across events in a scenario.	PSFs evolve across events in a scenario.	A sudden change in the scenario causes changes in the PSFs.
-	Levels of CPCs change continuously in driving process.	Driving maneuvers including braking, steering wheel operation, acceleration, indicator lights operation.

need to be generated to integrate dynamic features into these methods. Therefore in dynamic context, the static CREAM approach needs to be modified to evaluate human reliability.

In dynamic driving context, situations are continuously changing. As the context is assumed as the most important factor affecting human reliability in CREAM approach and the assumption of COCOM in CREAM that the control degree of human operators on context effects human performance [Hol98], dynamical human driver performance could be evaluated with new defined CPCs describing the dynamic features of situated driving context. As described in Table 4.2, with continuously changing levels of CPCs in driving process, the dynamic progression can be modeled. The dynamic initiator in driving maneuver are mainly braking, steering wheel operations, acceleration and deceleration, and the use of indicator lights operation as actions to express lane changing wishes and therefore affect the environment with stated intentions. So the new list of CPCs describing the dynamic driving context could be established.

4.4.1 New list of CPCs

In [HLLS21], the author established a list of CPCs to characterize the main elements affecting human drivers reliability in situated driving context. 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. These levels are defined by literature research and expert knowledge and could be used for unfuzzified HPRS calculation. The complete list of CPCs is shown in Table 4.3.

Table 4.3: New CPCs and related performance reliability for dynamic driving context [HS20]

CPC name	Level/Description		Expected effect on performance reliability
Number of surrounding vehicles (N)	N = 0		Improved
	$1 \leq N \leq 3$		Not significant
	$N \geq 4$		Reduced
Time to collision (TTC)	TTC > 5.5 s		Improved
	$2.5 \leq TTC \leq 5.5$ s		Not significant
	TTC < 2.5 s		Reduced
Ego-vehicle speed (V)	$V \leq 22$ m/s		Improved
	$22 < V \leq 30$ m/s		Not significant
	V > 30 m/s		Reduced
Longitudinal acceleration	$V \leq 22$ m/s	$a \leq 1.6$ m/s^2	Improved
		$1.60 < a \leq 2.32$ m/s^2	Not significant
		$a > 2.32$ m/s^2	Reduced
	$22 < V \leq 30$ m/s	$a \leq 1.13$ m/s^2	Improved
		$1.13 < a \leq 1.60$ m/s^2	Not significant
		$a > 1.60$ m/s^2	Reduced
	V > 30 m/s	$a \leq 1.13$ m/s^2	Not significant
		$a > 1.13$ m/s^2	Reduced
Lateral acceleration	$V \leq 22$ m/s	$a \leq 1.48$ m/s^2	Improved
		$1.48 < a \leq 2.15$ m/s^2	Not significant
		$a > 2.15$ m/s^2	Reduced
	$22 < V \leq 30$ m/s	$a \leq 1.05$ m/s^2	Improved
		$1.05 < a \leq 1.48$ m/s^2	Not significant
		$a > 1.48$ m/s^2	Reduced
	V > 30 m/s	$a \leq 1.05$ m/s^2	Not significant
		$a > 1.05$ m/s^2	Reduced
Traffic density	Low (≤ 7)		Improved
	Medium (8 - 14)		Not significant
	High (≥ 15)		Reduced
General visibility conditions	Daytime with sunny weather		Improved
	Early 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 [BR11], 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 [BB08]. 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 [VDHH93] 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 thesis, three levels of speed are identified, speed faster 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 [EMP16], 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 [EMP16], [ZB13]. 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 [EMP16], 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) [GKLB16], [SY05]. Considering identified TTC and ego-vehicle speed in this thesis, 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.

- 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.

A new list of CPCs based on Table 4.3 is generated to better describe driving situations, including ego-vehicle states (longitudinal speed, lateral speed, longitudinal acceleration, and lateral acceleration) and surrounding environment states (TTC front, TTC front left, TTC front right, TTC behind, TTC behind left, and TTC behind right). Here, totally 10 new CPCs are defined to illustrate the dynamic features of situated driving context and evaluate human performance reliability. It should be noted that the new CPCs (ego-vehicle speed and acceleration and the TTCs) are ego-vehicle centered.

4.4.2 Calculation of HPRS

The data clustering methods of FN-DBSCAN, CLUSTERDB*, and GMFPE as explained in chapter 4.3 are applied to automatic generation of the membership functions. The corresponding optimal parameters are defined with genetic algorithm. Therefore, the reliance of HRA methods is reduced and the results are just related to the data.

In the original CREAM approach, the CPC score is [$\sum$ reduced, $\sum$ improved], which denotes that the sum of reduced effects and sum of improved effects on performance reliability are considered as two features. Then the control mode can be defined with Figure. 4.1.

In this thesis, genetic algorithms and data clustering approaches are used as tools for ordering and classifying behaviors and situations, the membership functions will be assigned to different levels. The connection between the membership functions and assigned performance reliability is human driver's driving skills and experience. With the understanding that different group of clustered data indicates different driving behavior characteristics, the membership functions can be used to present the driving skills and experience as which directly reflect different driving behavior characteristics.

The membership function with improved effects can be assigned to the value of 1, with not significant effects to 0, and with reduced effects to -1. Then each CPC score can be calculated.

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.

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

$$HPRS = \lambda_1 \cdot \sum reduced + \lambda_2 \cdot \sum improved, \tag{4.18}$$

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 Figure 4.1, 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 Figure 4.1 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 Figure 4.2.

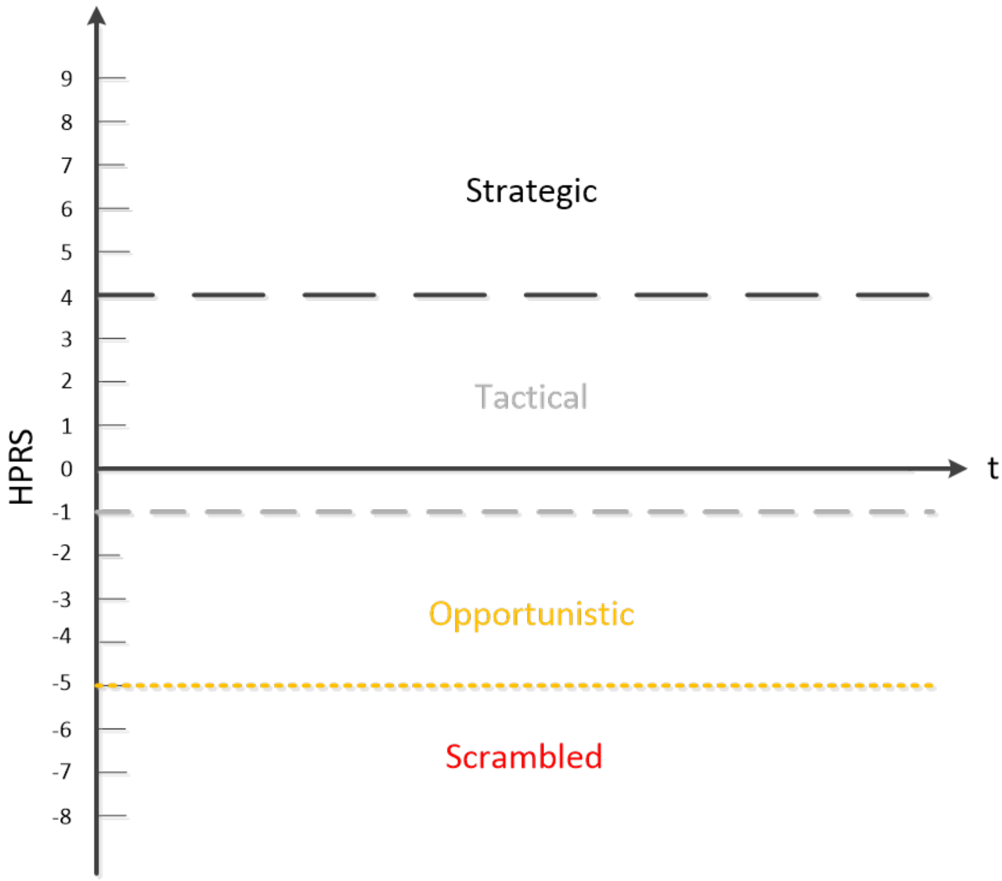

Figure 4.2: Relations between HPRS and control modes [HS20]

In general, the steps to obtain HPRS is shown in Figure 4.3. Firstly, the CPCs are defined to characterize the main factors describing the context. Next, the FN-DBSCAN, CLUSTERDB*, and GMFPE with genetic algorithm are executed to generate the membership functions. Furthermore, the obtained membership functions are assigned to different CPC levels to calculate each CPC score. Finally, all these ten CPCs scores are added up to obtain the HPRS.

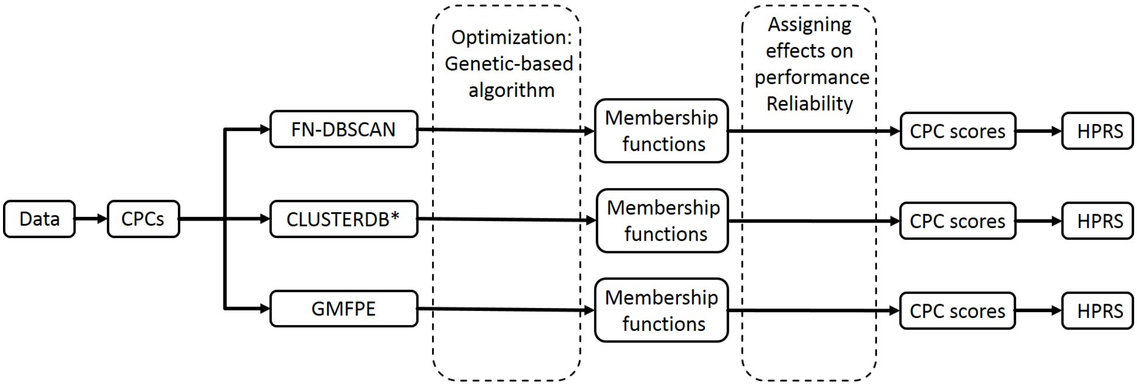

Figure 4.3: Flowchart to obtain HPRS

4.5 Summary

In this chapter, the modified fuzzy-based CREAM approach is established for the evaluation of human performance reliability in dynamic changing situations. This new approach can be applied for human reliability performance evaluation for dynamically changing complex situations in which continuously decisions and actions are realized within safety relevant context. The situated driving context is taken as an example for the application of this new and dynamic approach. In this case, the new CPC lists for unfuzzified HPRS and fuzzified HPRS depicting the main features of situated driving context are generated. Three data clustering approaches including FN-DBSCAN, CLUSTERDB*, and GMFPE are introduced for data clustering to define the cores and supports of membership functions, so the fuzzified and HPRS could be obtained, meanwhile, the reliance on expert knowledge in determining the CPC levels could be reduced.

5 Experimental results and analysis

In this chapter, driving simulator is introduced, membership functions with different data clustering approaches are presented. The final results with different membership functions are shown, and the HPRS results are compared and analyzed. The Swiss chess model is introduced to explain the critical behaviors detected by HPRS. An example is given to explain HPRS for situated and personalized monitoring of human driver behaviors.

Part of the contents, figures, and tables presented in this chapter are modified after previous publications [HS22][HBS22][HLLS21][HLL21][HS20][HTS20][He19]. Part of the contents, figures, and tables are prepared for publication of [He22a].

5.1 Data generation platform

A professional driving simulator SCANeR^{TM} studio as shown in Figure 5.1 is used to collect data. The simulator realize a 270° view of the driving environment, a rear view mirror, and two side mirrors. For controlling ego-vehicle, there is a base-fixed driver seat, steering wheel, and pedals are used. Data describing ego-vehicle

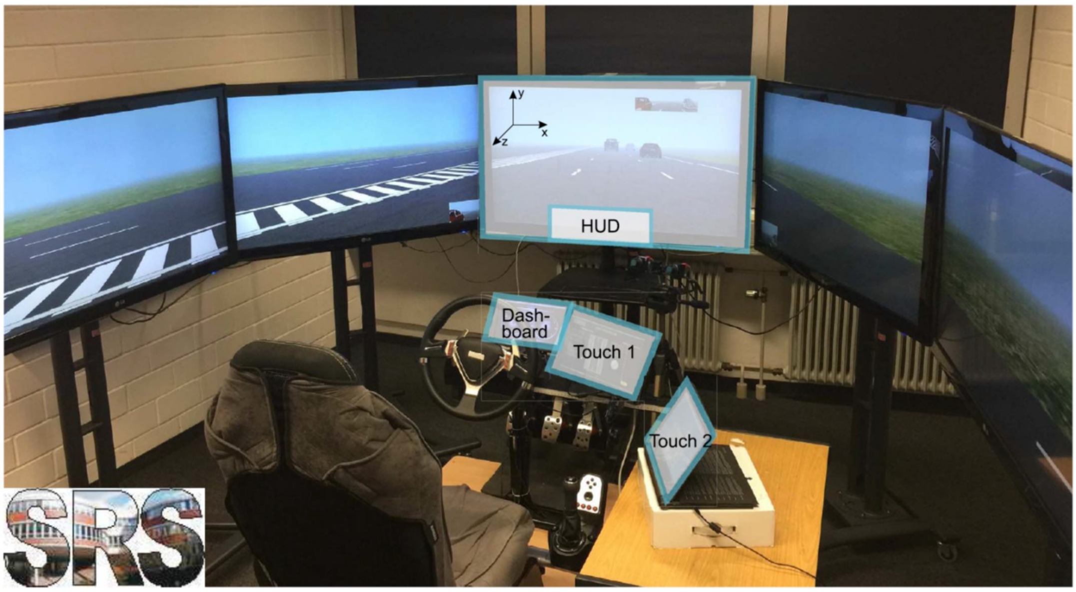

Figure 5.1: Driving simulator laboratory, Chair of Dynamics and Control, U DuE

dynamics (e.g. speed, steering angles, etc.) and surrounding interacting vehicle status (e.g. lateral shift, TTC, etc.) relative to ego-vehicle are collected allowing evaluation driver interaction behaviors also to be used for reliability analysis.

Driving scenarios are set on a three-lane dual carriage highway. Fog, curves, and undulations are introduced to generate the real driving environment. In addition to ego-vehicle, interacting vehicles are introduced to generate situated driving context which can continually stimulate ego-vehicle driver to perform various maneuvers. Driver may change lanes, decelerate, maintain relative speed as deemed appropriate in accordance with Germany's driving rules. Therefore, participants are required to drive on right lane unless overtaking or moving at approximately the same speed as other vehicles present in other lanes. With dynamically changing driving context and corresponding to change driving maneuvers, driver's reliability varies over time.

Roads in highway scenario have their own characteristics differing from other roads, like urban roads. Highway roads are usually in closed road design, wide, flat, and less changed road conditions. These features could induce some driving issues that differ from other roads. 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 120 km/h is allowed in highway scenario, but it is not allowed in urban roads. To increase the complexity during manual driving process, drivers are asked to exist the highway and then return back. With the dynamically changing situations in driving, drivers' reliability will also fluctuate, which can be evaluated online by the proposed method.

5.2 Unfuzzified HPRS results

5.2.1 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 Figure 5.2.

From Figure 5.2, 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.

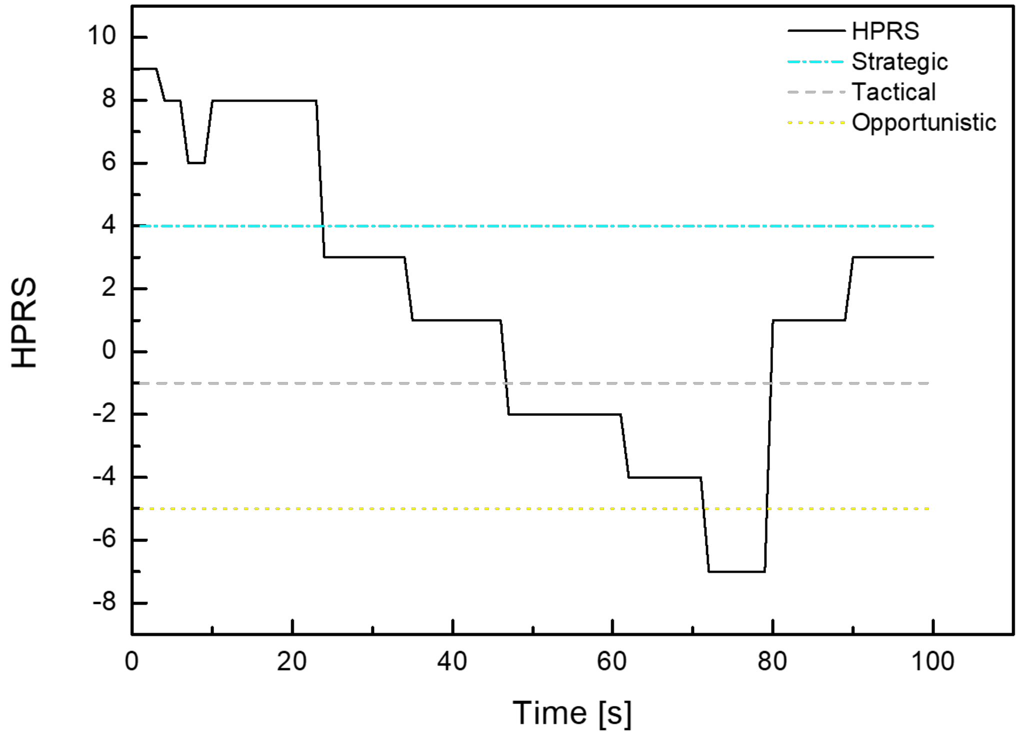

Figure 5.2: Case study of artificial HPRS [HTS20]

5.2.2 Experimental results

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 results of HPRS with the example scenario is shown as Figure 5.3. It could be obtained that the unfuzzified HPRS values jumps between discrete numbers and all HPRS values are all above the tactical level.

5.3 Membership function results with different approaches

Two participants with driving license for more than 10 years are involved in the experiment with the same scenario. Driving scenario are set on a three-lane dual carriage highway. In addition to ego-vehicle, interacting vehicles are introduced to generate situated driving context which can continually stimulate ego-vehicle driver to execute various maneuvers. As the driver possess different levels of skills and experience on different situations, the human performance reliability of the driver on situations is also varying. The driving data with 120 s for each participant are selected to generate the membership functions and HPRS.

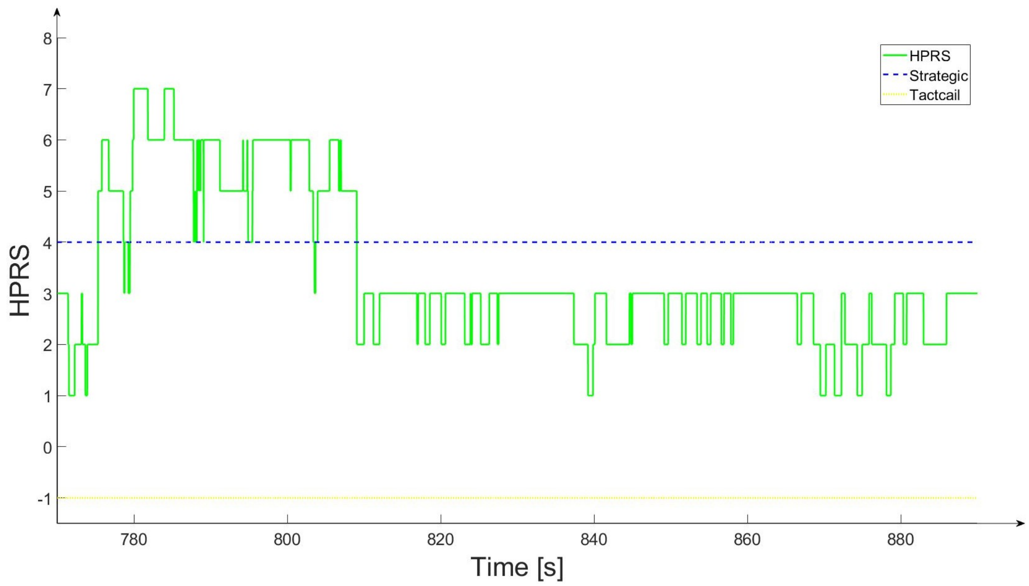

Figure 5.3: Unfuzzified HPRS of example scenario [HTS20]

The membership functions of ten CPCs are determined with the three described data clustering approach for each participant. The CPCs of the surrounding environment states (e.g. six TTC related CPCs) are grouped to be applied for data clustering as the nature of characteristics in TTC-related crisp data are the same. In these case, the membership functions of TTC are generated. The membership functions of participant_1 with different data clustering approaches are shown in Figure 5.4, Figure 5.5, and Figure 5.6. The membership functions of participant_2 with different data clustering approaches are shown in Figure 5.7, Figure 5.8, and Figure 5.9. It could be detected that different number of membership functions are obtained for each CPCs with different clustering approach. The GMFPE obtains three membership functions for each CPC as the predefined number of membership functions is set to be three to match the performance reliability levels in CREAM approach. The obtained number of membership functions from FN-DBSCAN and CLUSTERDB* is varying from 1 to 3. In this case, membership functions should be assigned to the expected effect levels logically and considering the actual situation. For example, membership function with low speed should be assigned with improved effect on performance reliability and high speed with reduced effect. For TTC, it is opposite as small TTC should be assigned with reduced effect and large TTC with improved effect. When only one membership function is obtained (membership function of lateral speed with FN-DBSCAN of participant_1), three segments in this membership function should be assigned to different effect levels. Here the segment with membership degree of 1 is assigned to not significant effect, other segments are assigned to reduced effect.

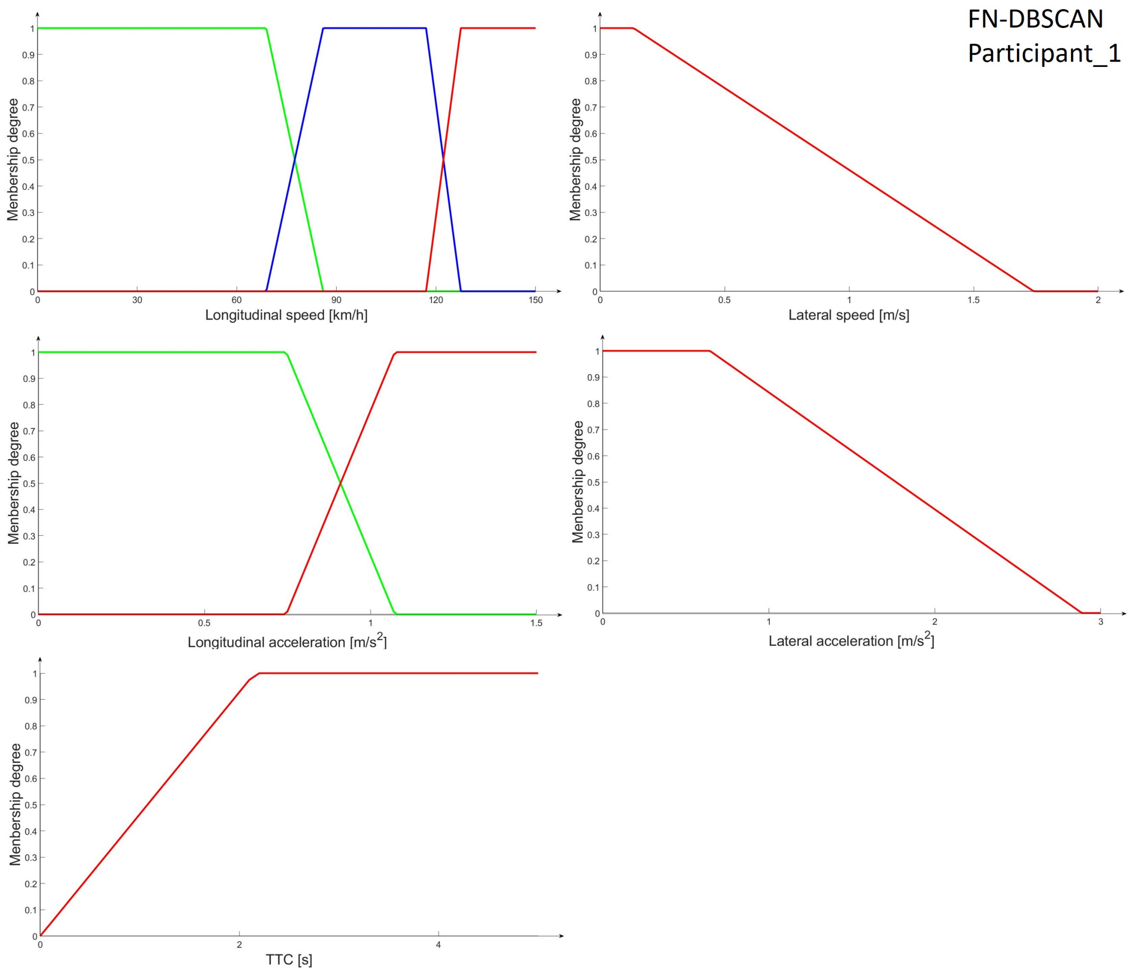

Figure 5.4: Membership functions of participant_1 with FN-DBSCAN approach [He22a]

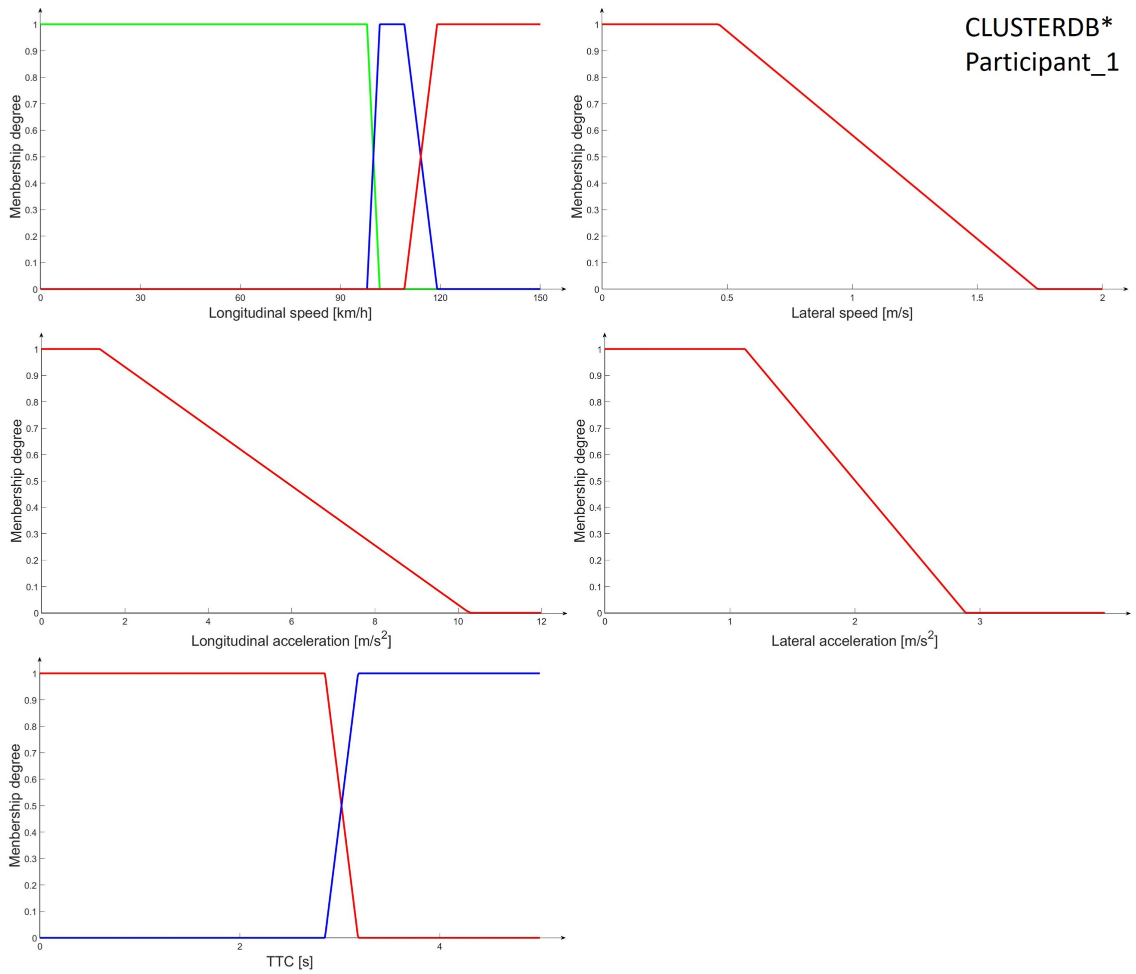

Figure 5.5: Membership functions of participant_1 with CLUSTERDB* approach [He22a]

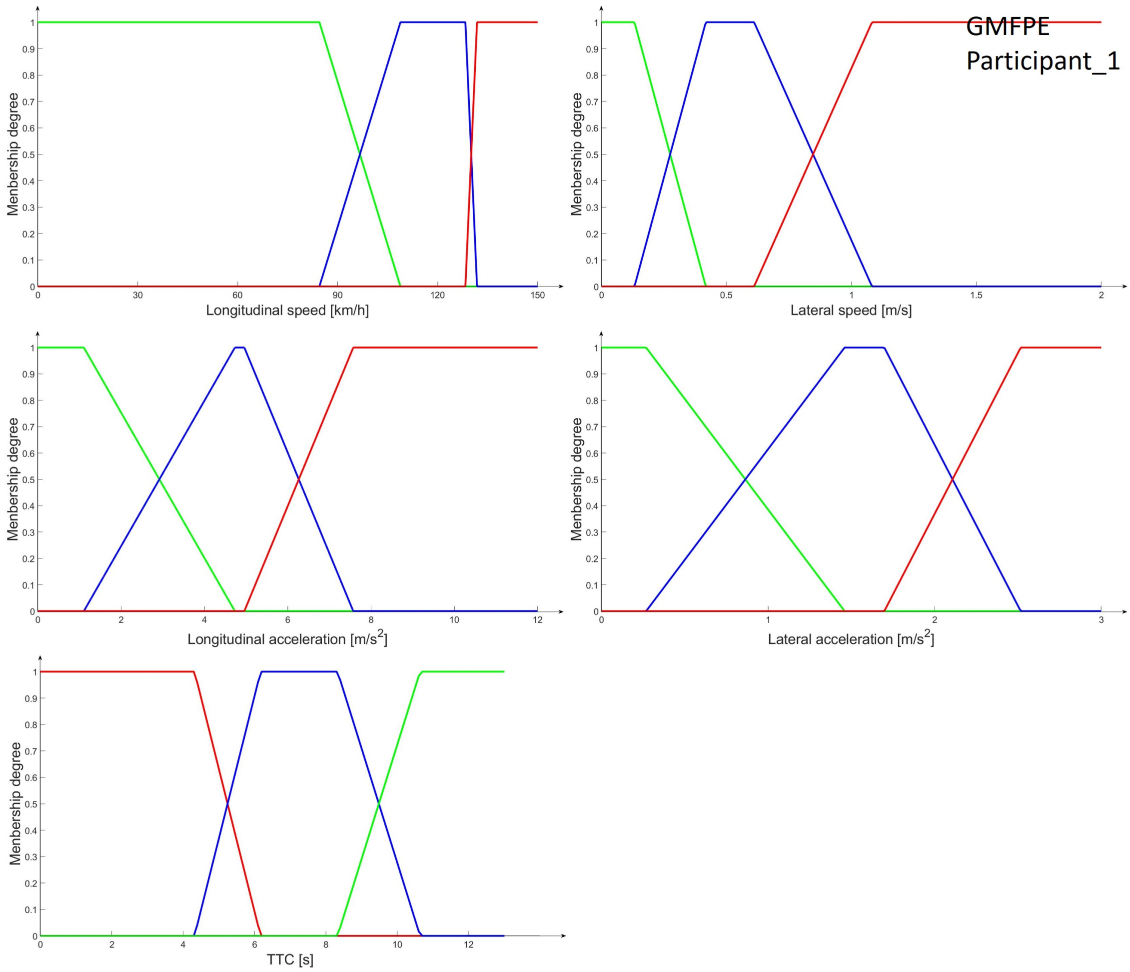

Figure 5.6: Membership functions of participant_1 with GMFPE approach [He22a]

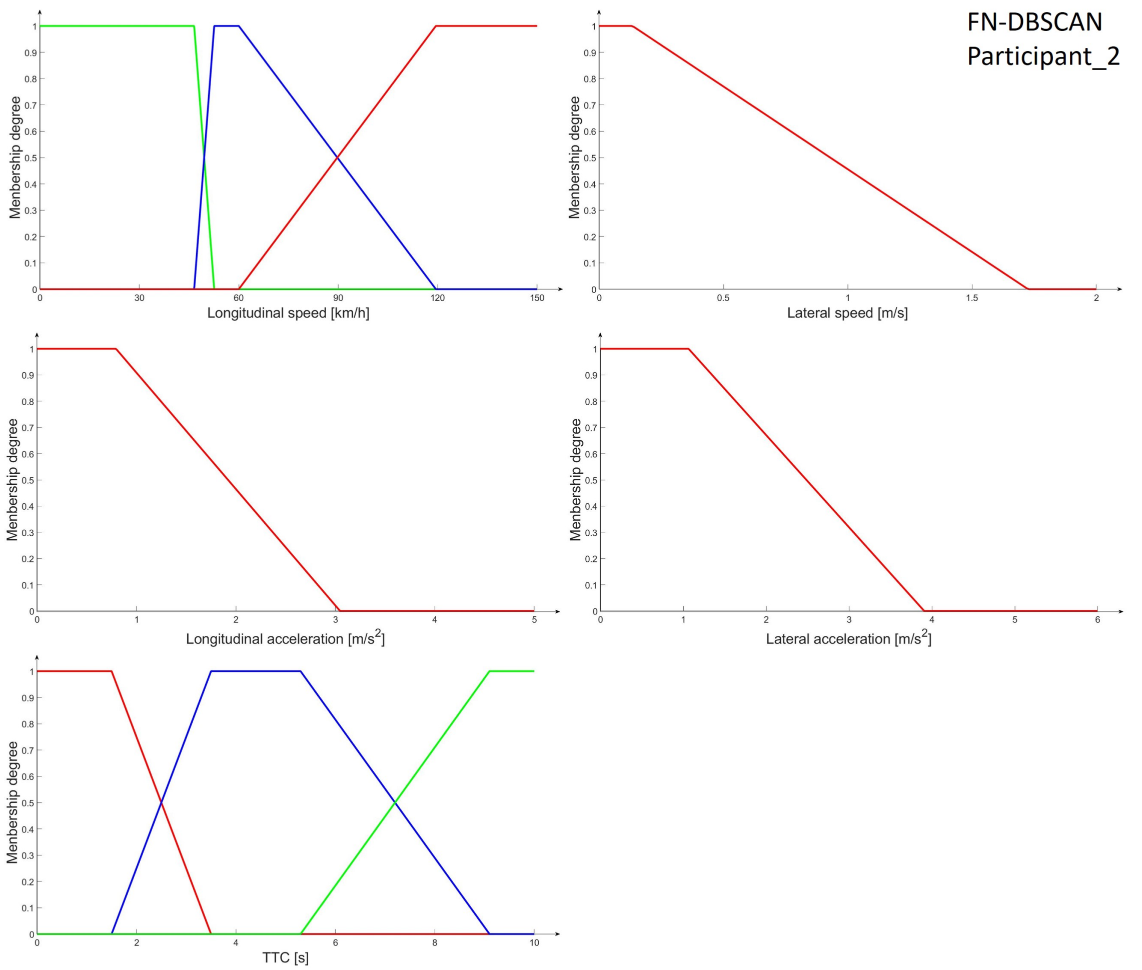

Figure 5.7: Membership functions of participant_2 with FN-DBSCAN approach [He22a]

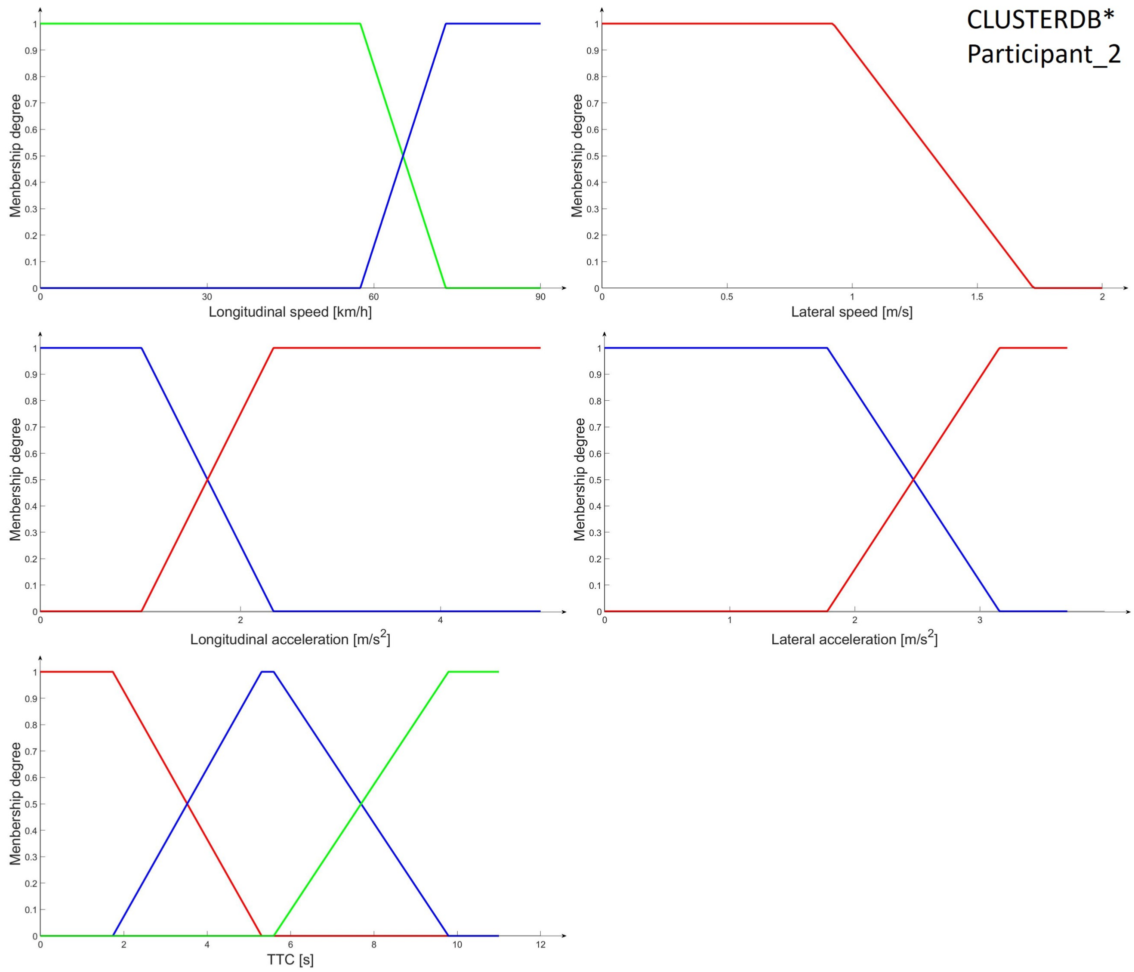

Figure 5.8: Membership functions of participant_2 with CLUSTERDB* approach [He22a]

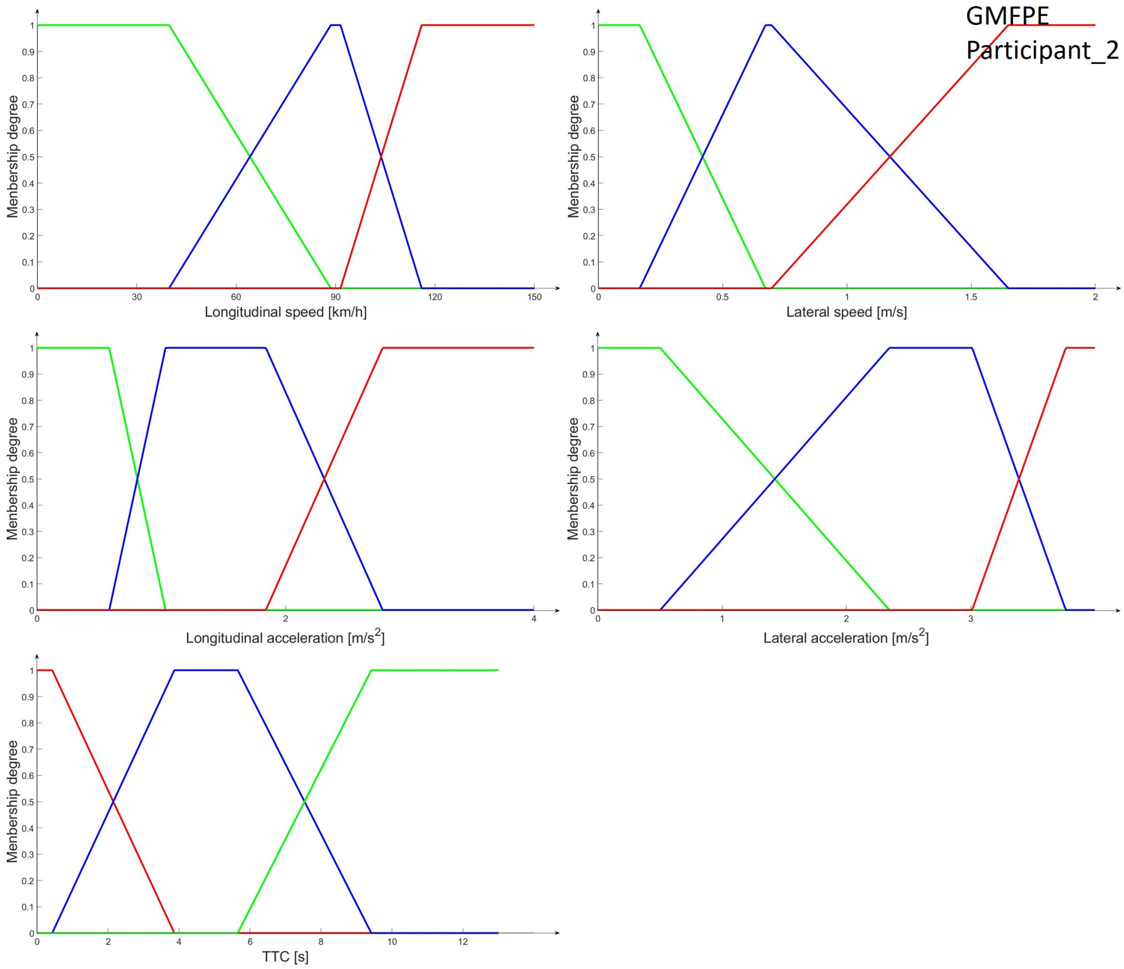

Figure 5.9: Membership functions of participant_2 with GMFPE approach [He22a]

5.4 Human performance reliability score (HPRS) with different approaches

When all membership functions are assigned to different effect levels, the CPC score for each CPC could be calculated. The CPC score results for participant_1 are shown in Figure 5.10, Figure 5.11, and Figure 5.12. The CPC score results for participant_2 are presented in Figure 5.13, Figure 5.14, and Figure 5.15. It could be concluded that the results for different clustering approaches are different, especially the CPC scores obtained by FN-DBSCAN and CLUSTERDB* are varying from -1 to 0 except the CPC score of longitudinal speed, but the CPC scores obtained by GMFPE are varying from -1 to 1.

When all CPC scores are calculated, the final HPRS of two participants with different clustering approach could be determined, which is the sum of each CPC score with respect to time, as shown in Figure 5.16 and Figure 5.17. It could be observed that HPRS fluctuate with time continuously which indicates the performance reliability of the participants varies with different situations, as well as different effects affecting the values.

For participant_1, HPRS from three clustering methods are above opportunistic level, and HPRS from GMFPE are above tactical level. In general, the HPRS values from FN-DBSCAN and CLUSTERDB* methods do not differ much and fluctuate around the tactical level. The HPRS from GMFPE results fluctuate largely above the strategic level, and only for some periods of time do the values fluctuate below the strategic level. The reason for the difference between HPRS values from GMFPE and values from other two methods is mainly due to the difference in assigning effect levels to membership functions. In membership functions obtained by FN-DBSCAN and CLUSTERDB*, the effect levels are assigned with not significant effects and reduced effects when only one or two membership functions obtained for each CPC, so the CPC scores are varying between -1 and 0. However, the effect levels of membership functions obtained by GMFPE could be assigned with improved, not significant and reduced effects as the obtained number of membership functions in GMFPE is predefined as three, so the CPC scores are varying between -1 and 1. Especially for the membership functions of TTC as which affects the CPC scores of six TTC related CPCs. For example, only one membership function is obtained with FN-DBSCAN method, so the sloping segment is assigned with reduced effect and the horizontal segment with not significant effect. In this case, the CPC scores of TTC with FN-DBSCAN are varying from -1 to 0. The TTC data obtain three membership functions using GMFPE, so they could be assigned to three different effect levels, which result in the CPC scores varying between -1 and 1.

The HPRS results with three clustering methods from participant_2 are above tactical level and fluctuating near strategic level, while only for a few times, the results fall below the strategic level. It could be concluded that unlike the HPRS results of

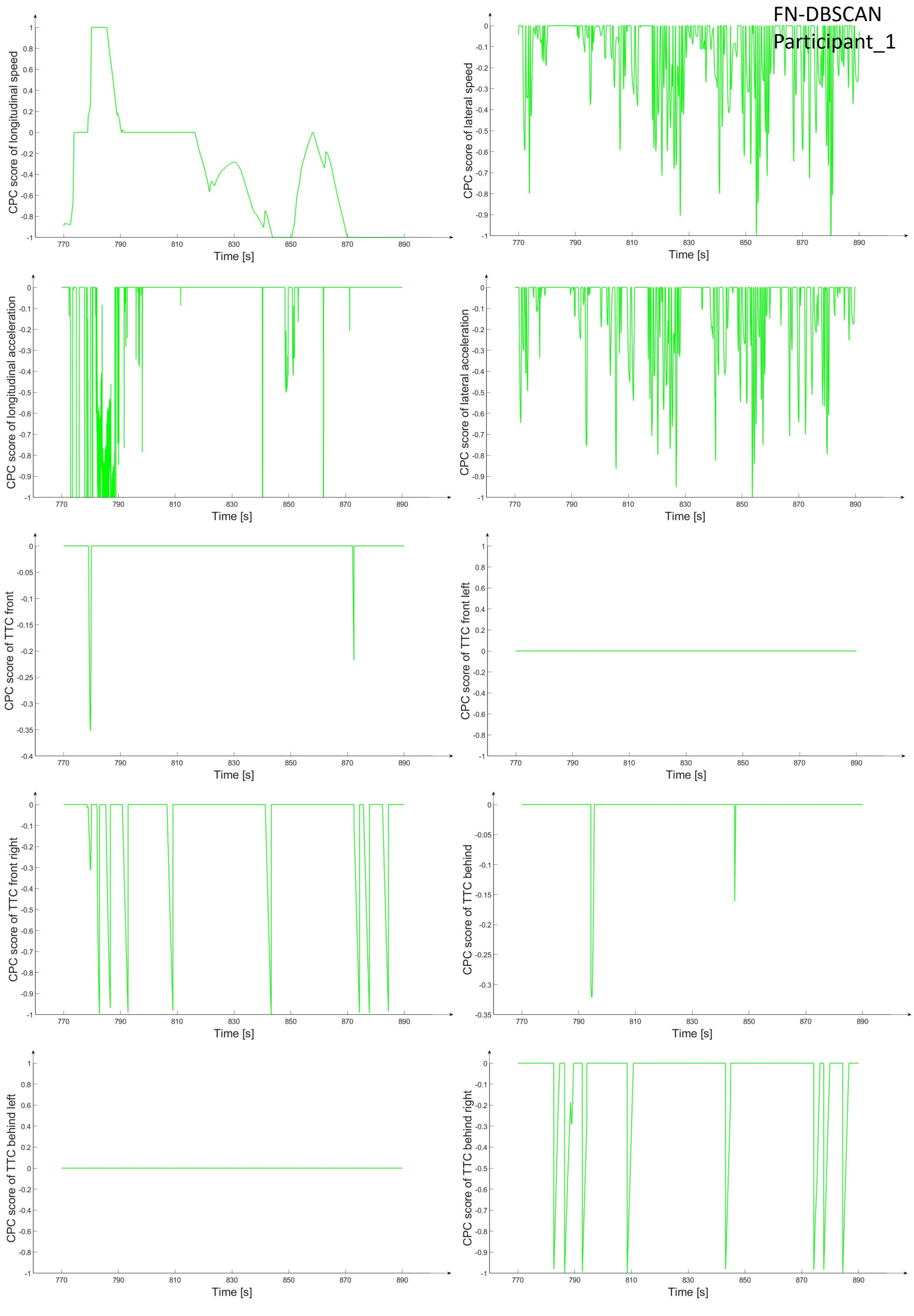

Figure 5.10: CPC score of participant_1 with FN-DBSCAN data clustering approach [He22a]

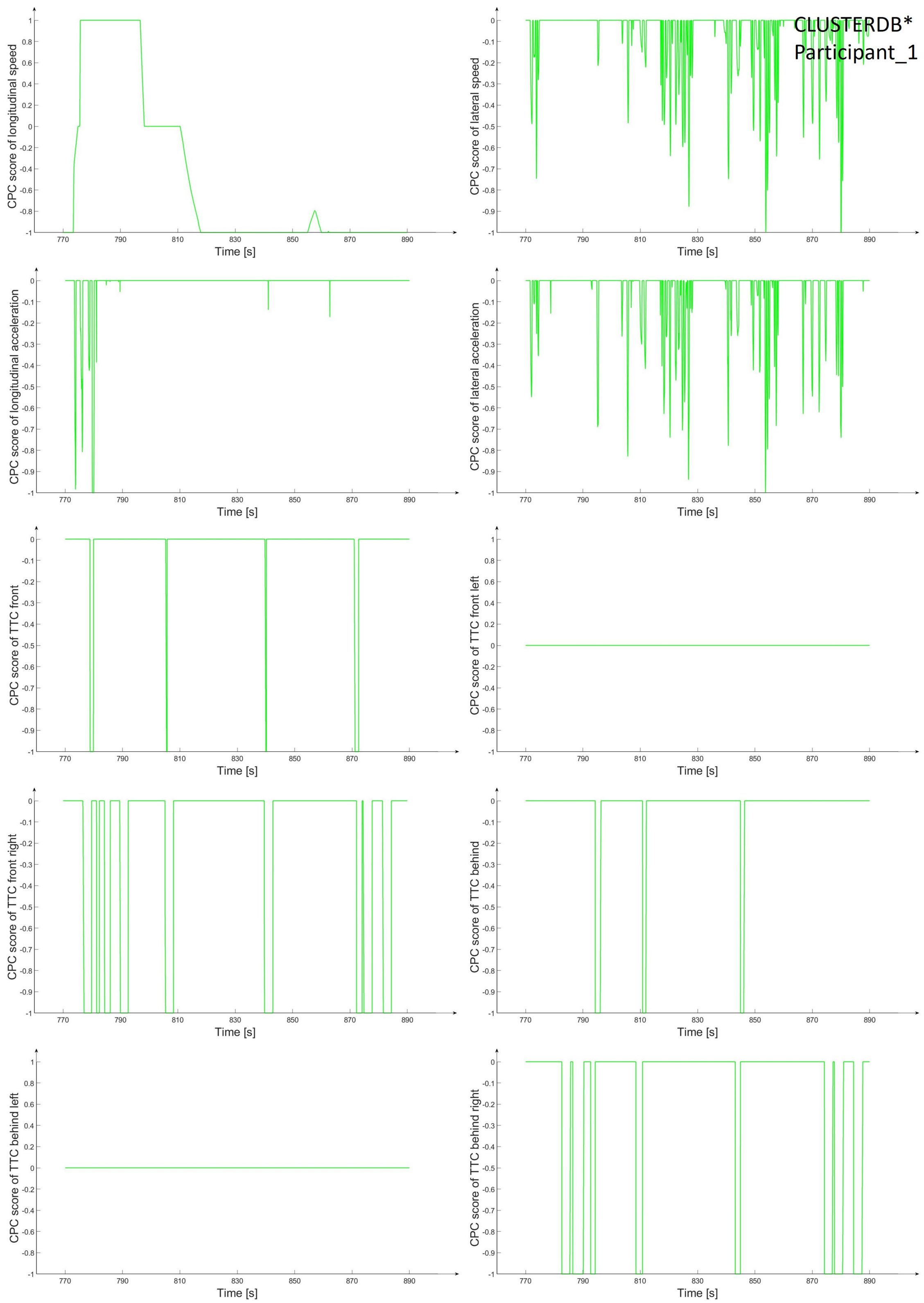

Figure 5.11: CPC score of participant_1 with CLUSTERDB* data clustering approach [He22a]

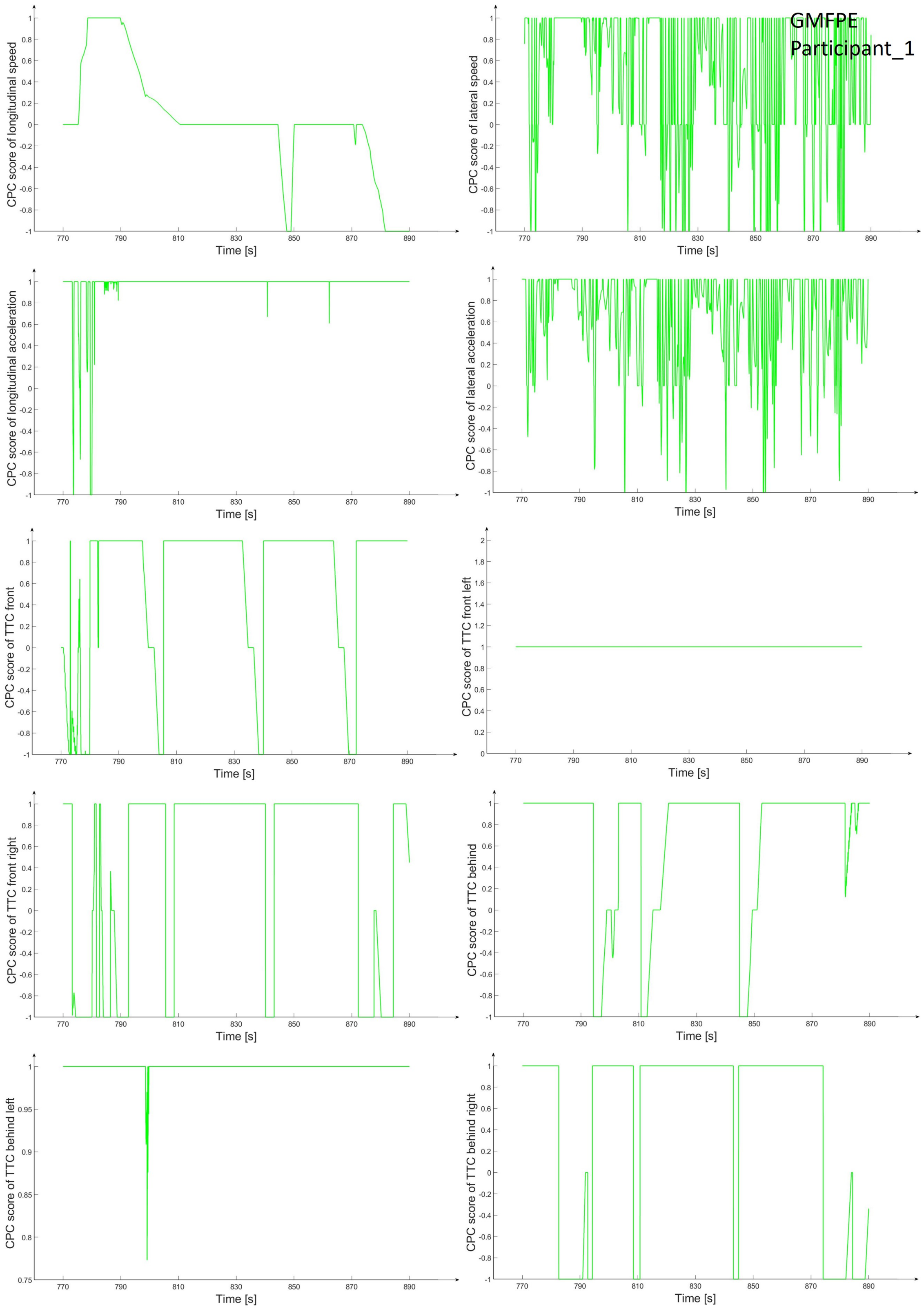

Figure 5.12: CPC score of participant_1 with GMFPE data clustering approach [He22a]

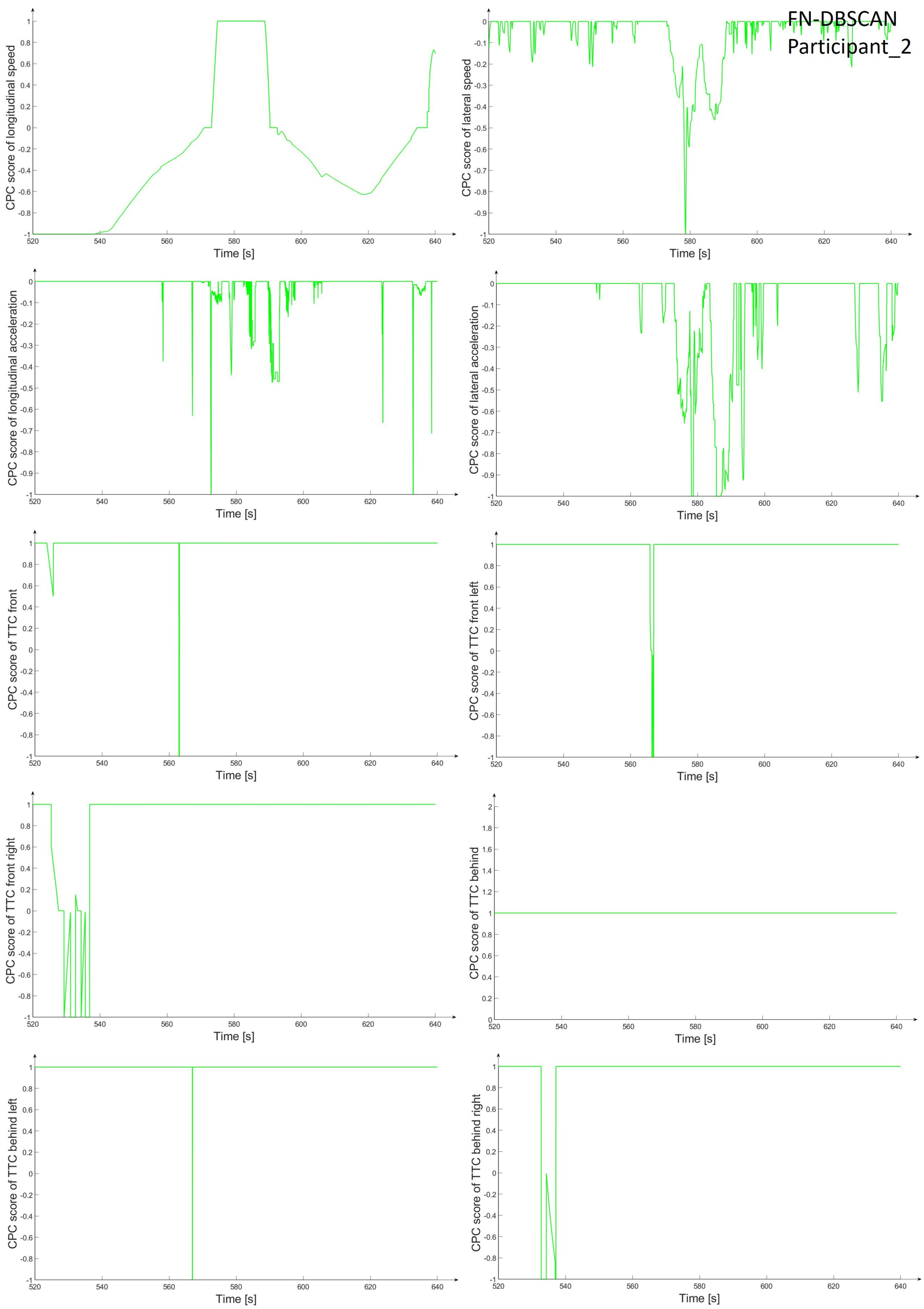

Figure 5.13: CPC score of participant_2 with FN-DBSCAN data clustering approach [He22a]

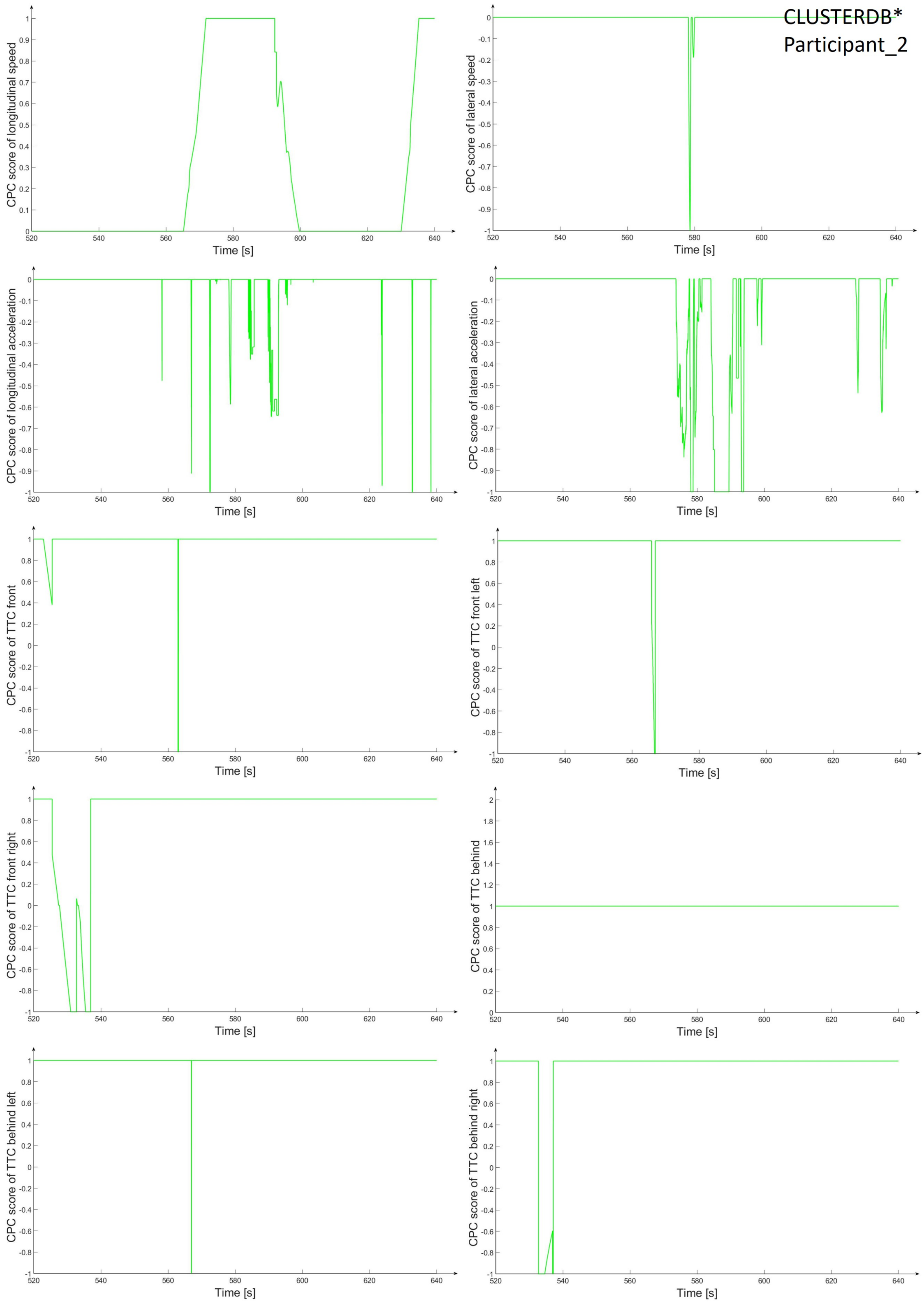

Figure 5.14: CPC score of participant_2 with CLUSTERDB* data clustering approach [He22a]

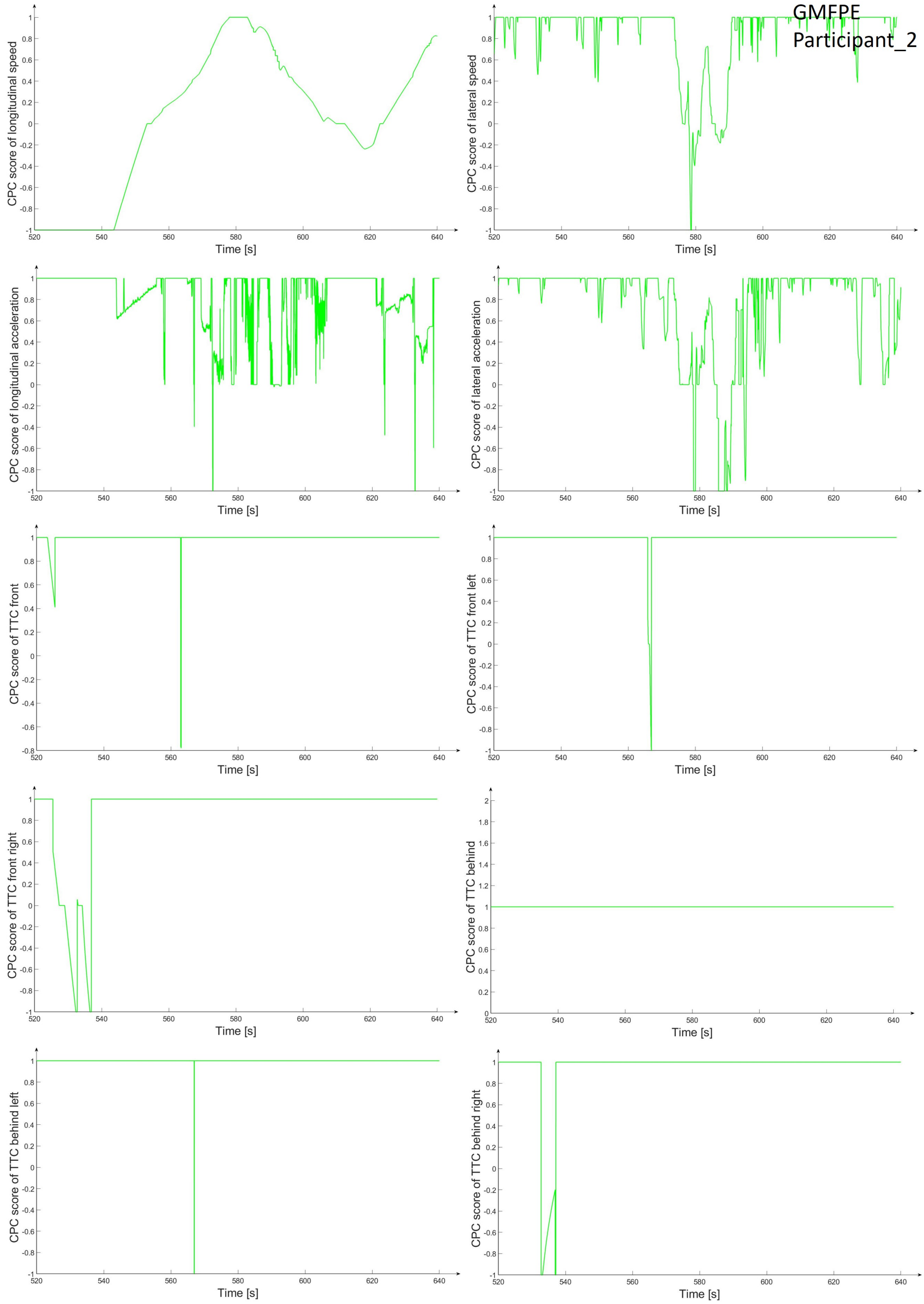

Figure 5.15: CPC score of participant_2 with GMFPE data clustering approach [He22a]

GMFPE for participant_1, which differ significantly from the remaining two methods, the HPRS results of three methods for participant_2 are more or less similar, although the results of GMFPE are still above the results from other two methods.

Therefore, membership functions obtained from FN-DBSCAN and CLUSTERDB* could more detailed/realistically characterize the personal driving behaviors as driving data are clustered based on the characteristics of the data themselves. So in this case, the HPRS results from FN-DBSCAN and CLUSTERDB* give a more accurate picture of human performance in situated driving process.

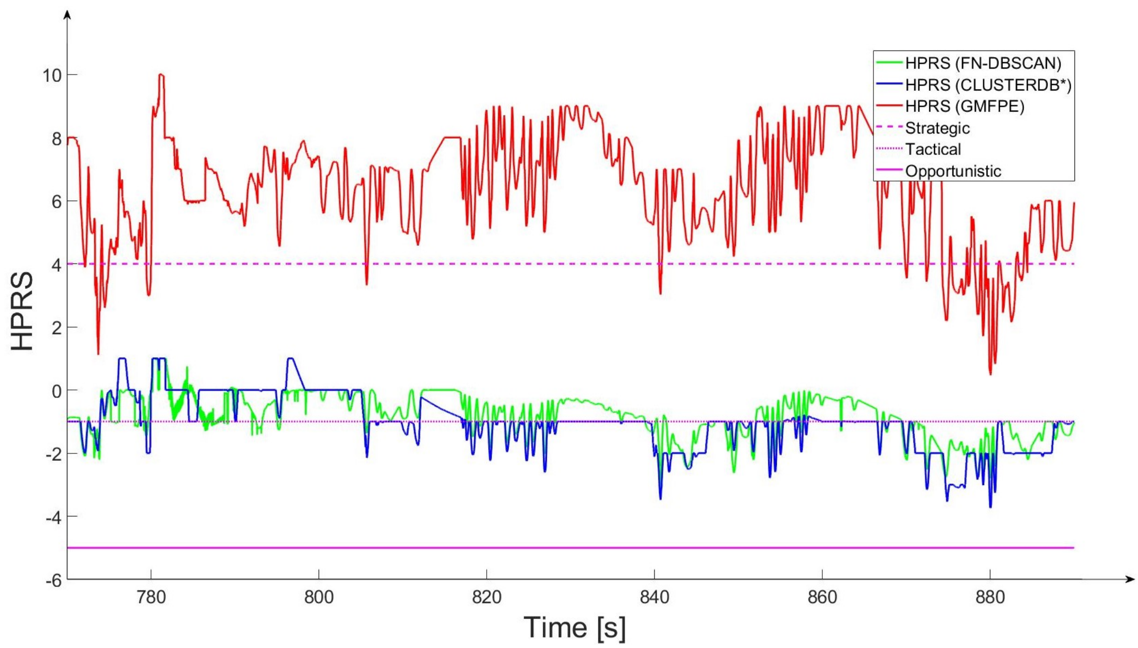

Figure 5.16: HPRS results of participant_1 with different data clustering approach [He22a]

5.5 HEP intervals transforming between CREAM and SRK model

The connection between CREAM approach and SRK model can be established with the comparison of HEP values in Table 4.1 and Table 3.7. It could be obtained that the HEP intervals of the control modes in CREAM overlap with the HEP intervals of the behavior levels in SRK mode, which could be intuitively detected from Figure 5.18. With the HEP intervals and overlapping of CREAM and SRK model, the new HEP intervals of SRK model for the connection with HPRS can be generated. For example, the minimum HEP values of strategic mode and skill-based level are 0.00005 and 0.00007 respectively, so the new minimum HEP value for skill-based

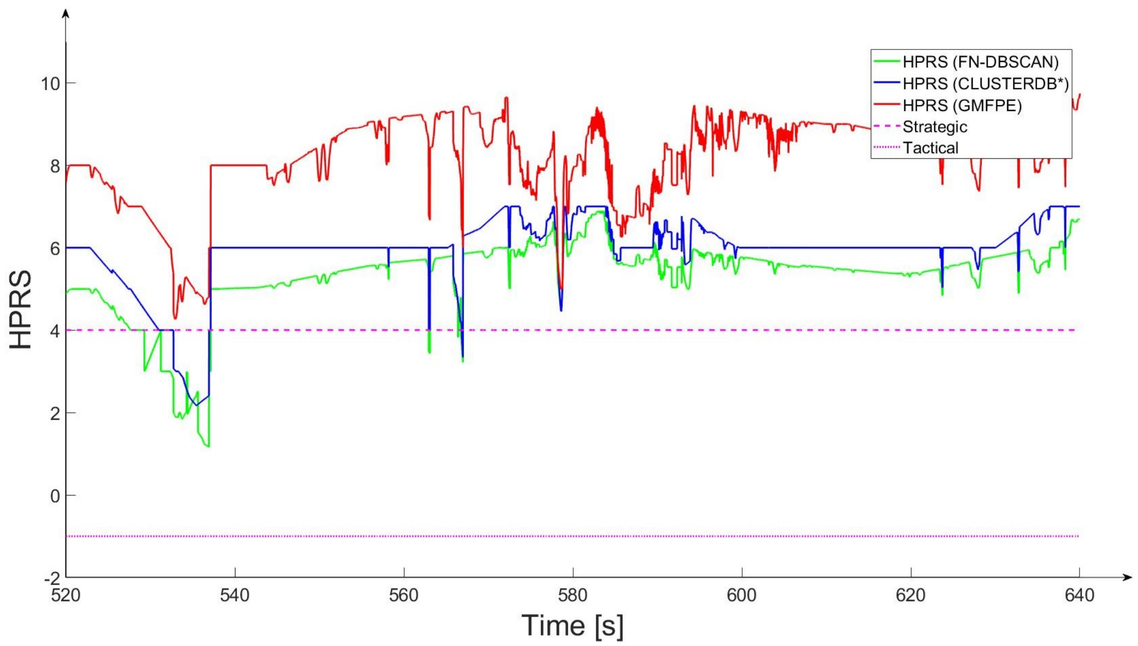

Figure 5.17: HPRS results of participant_2 with different data clustering approach [He22a]

level could be 0.00007. The maximum values of strategic mode and skill-based level are 0.01 and 0.0053 respectively, so the new maximum HEP value for skill-based level could be defined as 0.005. The new HEP intervals for SRK model are presented in Figure 5.19.

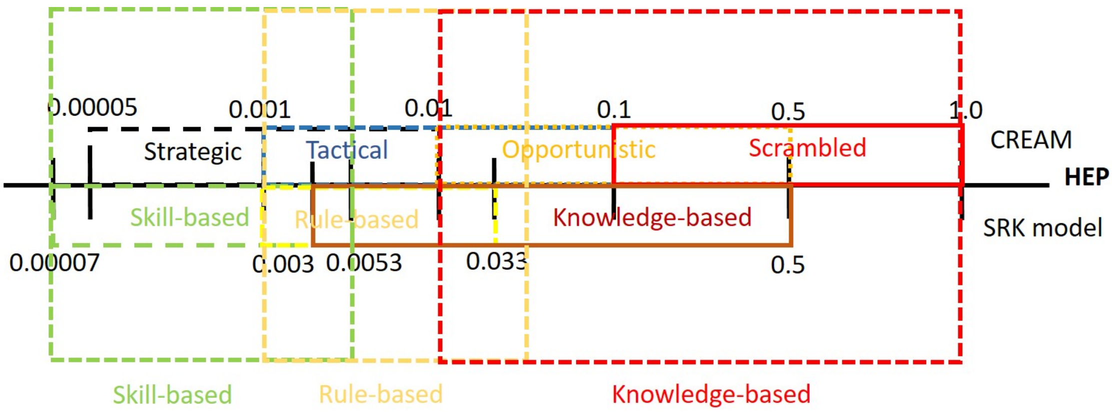

Figure 5.18: The HEP intervals transition between CREAM and SRK model

When the connection between control modes and behavior levels are defined, the structure of HPRS with control modes in Figure 4.2 can be converted into the structure of HPRS with behavior levels, shown in Figure 5.20. From Figure 5.18,

Control modes	Behavior levels	HEP interval
Strategic mode	Skill-based	(0.00007, 0.005)
Tactical mode	Rule-based	(0.001, 0.05)
Opportunistic mode		
Scrambled mode	Knowledge-based	(0.01, 1.0)

Figure 5.19: The HEP intervals of behavior levels for the connection with HPRS

the left boundary of opportunistic mode is 0.01, which is consisted with the left boundary of knowledge level, so the HPRS value of knowledge-based level can be defined as the same with HPRS of opportunistic mode. The HPRS value of skill-based level is determined as 3 with the consideration of the right boundary of skill-based level of HEP value is slightly less than the right boundary of tactical mode of HEP value. With the SRK related HPRS framework, the HPRS results obtained by the modified fuzzy-based CREAM approach could be evaluated, which are shown in Figure 5.21 and Figure 5.22. For participant_1, the HPRS results calculated with FN-DBSCAN and CLUSTERDB* are fluctuating around the rule-based level and some of the results are continuously in knowledge-based level, indicating human driver performance reliability is not optimal and human experience is not enough for the situations. However, the HPRS results obtained with GMFPE are almost above skill-based level. The reason for the significant differences between GMFPE result and other two results has been explained in chapter 5.4. For participant_2, all HPRS results are above rule-based level and most of the results are in skill-based level, indicating human performance reliability is optimal.

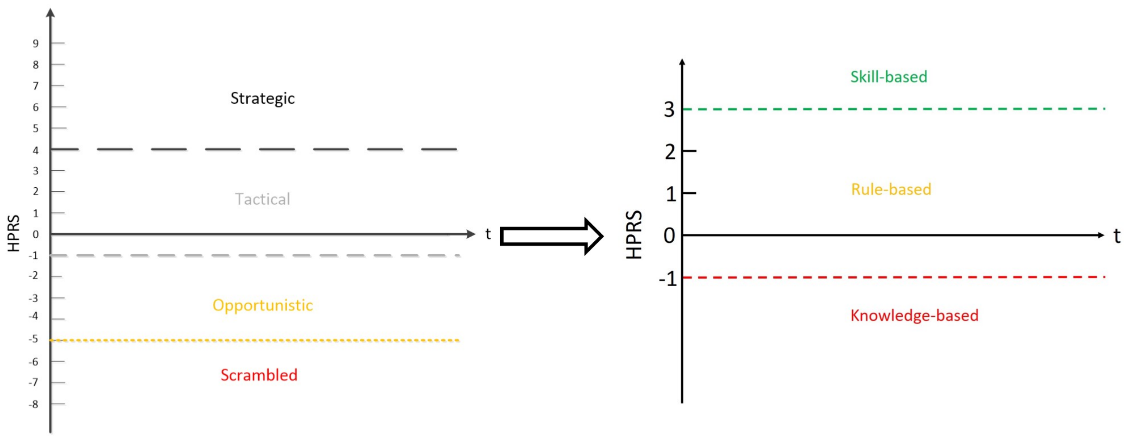

Figure 5.20: Transition of CREAM related HPRS to SRK related HPRS

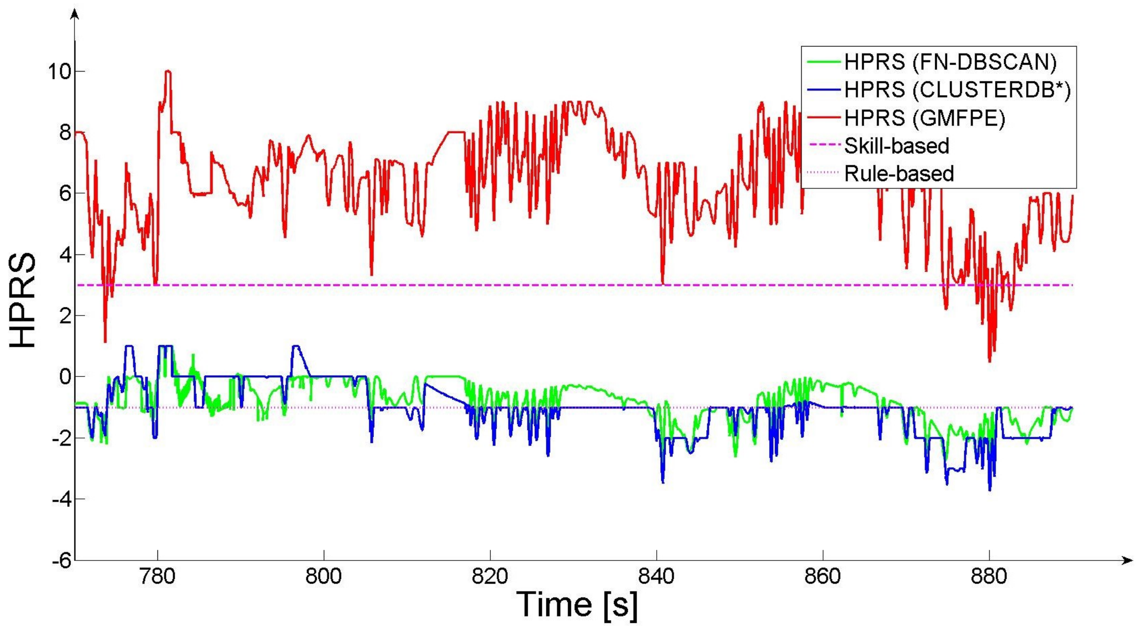

Figure 5.21: HPRS results of participant_1 with the evaluation of SRK levels

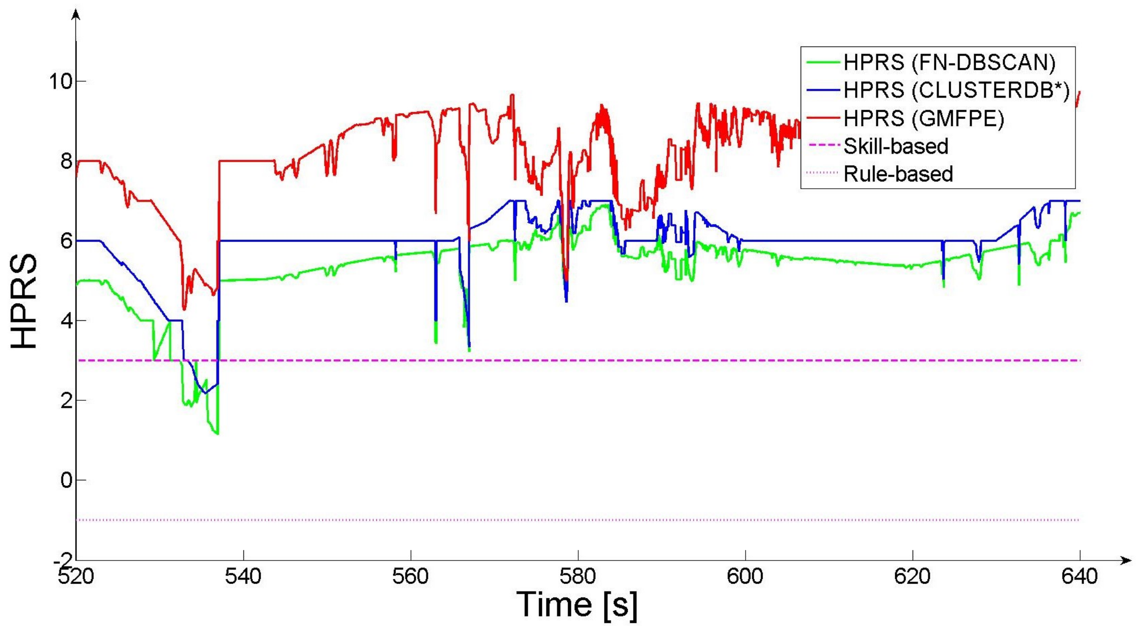

Figure 5.22: HPRS results of participant_2 with the evaluation of SRK levels

5.6 Discussion

Human-machine interaction and human performance reliability in dynamic context can be related to each other. Further the different human operators experiences and skills in general and for specific situations can be considered in this mapping. changing continuously as human operators possess different experience and skills on different situations. The traditional HRA approaches are considered as static, here the states of the PSFs and the influence of the PSFs are not changing over time. When the evolution of PSFs with time are accounted for, it becomes to the dynamic human reliability which needs to be evaluated with a new and dynamic HRA method capable of capturing the main features of dynamic context. To realize such a dynamic approach in this thesis the well-known CREAM approach is modified so that a dynamic fuzzy-based CREAM approach is established for the evaluation of human performance reliability. Situated driving context is taken as an example to demonstrate the evaluation process. A new concept of HPRS is proposed to define the human performance reliability quantitatively.

5.6.1 The features of the new approach

The new approach is established with the idea that human performance reliability is assumed as dynamic in situated context. In this case, the CREAM approach is modified to characterize the dynamic context and situated driving context is taken as an example to explain how this method works. In Figure 5.3, the unfuzzified HPRS results are obtained with CPC levels defining by literature research and expert knowledge. Comparing with the fuzzified HPRS result, it is obtained that unfuzzified HPRS results are non-smooth and therefore jumping between discrete integers, while the fuzzified HPRS numbers fluctuate continuously and more smoothly indicating that fuzzified results present more realistic changes in human performance reliability.

Different data clustering approaches are applied to driving data to determine the cores and supports values of fuzzy membership functions of CPCs. Therefore, the reliance on expert knowledge to determine the parameter values of membership function is reduced. At the same time, the reliance on expert knowledge to define the CPC levels in CREAM approach is also reduced as the CPC levels in Table 4.3 are determined with literature research and expert knowledge. With the clustering approaches, the CPC levels are determined by characteristics of the driving data itself.

Human drivers possess different driving habits and experience and skills with situations. When their driving data are analyzed and evaluated with clustering approaches individually, their individualized performance could be characterized with the obtained membership functions. The calculated HPRS is finally used for the quantitatively evaluation of human performance reliability in dynamic context in real time.

Briefly formatted, the main feature of the modified fuzzy-based CREAM approach compared with the existing HRA methods are:

- Fuzzified HPRS results indicating more realistic fluctuation of human performance reliability in situated driving context.
- Reduction of the reliance on expert knowledge to determine the CPC levels in CREAM approach and parameter values in membership functions of CPCs.
- Characterization of individualized performance with data clustering approaches.
- Establishing of a new concept to model the changing human reliability performance related to changing situated tasks and complexity.
- Quantitatively evaluation of human performance reliability in dynamic context in real time.

5.6.2 Explanation of HPRS results

It could be detected from Figure 5.16 and Figure 5.17 that HPRS values fluctuate (between different control modes), which indicates that the human reliability varies with time! Especially for the HPRS results from FN-DBSCAN and CLUSTERDB* of participant_1 comparing with the HPRS results from participant_2, the results from participant_1 are fluctuating near the tactical level, while the results from participant_2 are mainly above the strategic level, representing the human performance reliability of participant_1 during the analyzed period of time is worse than the performance of participant_2.

Human critical behaviors are always the concern of HRA as which weaken the safety and resilience of the whole human-machine system. Therefore, the identification of human critical behaviors or human errors from a more fundamental perspective, is essential for the evaluation of the influences and consequences of errors leading and the reduction of the occurrence frequency of errors [Phi18]. Many human error models have been developed in the process of increasing the understanding of human error. The wildly used model is Swiss chess model.

The Swiss chess model established by [Rea00] is a graphical method to conceptualize organizational accidents with different barriers on the one hand or with the understanding of a rare combination of circumstances for accidents on the other hand in many productive systems [LLC20]. This method consists of defenses/barriers/safeguards and weaknesses on the defensive layers. It is similar with the Swiss chess when the barriers in a system are represented by slices of chess, and the weaknesses are explained by holes. The presence of holes in slices does not normally lead to a critical outcome. However, when the holes in many slices momentarily line up to permit a trajectory of accident opportunity, critical outcomes will happen [Rea00]. In [Rea00], the holes are induced by two reasons: active failures and latent conditions. The active failures are the unsafe acts committed by operators including slips, lapses, mistakes, and violations. The latent conditions are indirect state that negatively affect the system, which can translate into error provoking conditions such as time pressure, fatigue, and inexperience and they can generate long-term holes like untrustworthy alarms and unworkable procedures.

As the Swiss chess model is applied to organization and system and slices are defined with factors such as organization, environment, and individuals, when it is used for the analysis of human driver critical behaviors, the Swiss chess model could be modified with two analysis directions: one is the typical latent and active failures structure, the other one only extracts the slice-hole framework from Swiss chess model, meanwhile, defines slices as CPCs in modified fuzzy-based CREAM approach and holes as experience and skills human driver owned.

For the first direction, two layers containing latent conditions and active behaviors could be determined for the analysis of human driver behaviors in situated driving context. In the latent conditions layer, human factors affecting human driving performance should be considered, such as time pressure, experience, fatigue, vigilance, etc. The active behaviors layer contains the error types based on driver error taxonomies. In this thesis, a generic driving error taxonomy is used, including action errors, cognitive and decision making errors, observation errors, information retrieval errors and violations [SS09]. In this previously introduced example, the HPRS results in the time range from 870 s to 880 s with FN-DBSCAN of participant_1 are selected as an example to demonstrate the analysis of human driver critical behaviors with Swiss chess model. This part of HPRS is selected because of its frequently fluctuation under tactical level. The situations could be detailed as follows:

- At time 870 s - 873 s, the ego-vehicle is braked (longitudinal speed $\approx$ 127 km/h, longitudinal acceleration $\leq$ 0.78 m/s^2) and changed to the left lane (lateral acceleration $\leq$ 2.23 m/s^2 to left direction, lateral speed $\leq$ 0.36 m/s) as there are a vehicle in front (TTC_front $\geq$ 1.69 s) and a vehicle in front right (TTC_front_right $\geq$ 1.32 s).
- At time 873 s - 875 s, the ego-vehicle has been changed to left lane (lateral acceleration $\leq$ 1.79 m/s^2 to right direction, lateral speed $\leq$ 0.24 m/s), the gas pedal is slightly pressed (longitudinal speed $\approx$ 128 km/h, longitudinal acceleration $\leq$ 0.21 m/s^2) and the ego-vehicle is passing right vehicle (TTC_front_right$\leq$ 3.49 s, TTC_behind_right$\leq$ 12 s).

- At time 875 s - 880 s, the ego-vehicle is roughly maintained the speed (longitudinal speed $\approx$ 129 km/h, longitudinal acceleration $\leq$ 0.42 m/s^2), the steering wheel is frequently turned left and right (lateral acceleration $\leq$ 2.42 m/s^2, lateral speed $\leq$ 0.5 m/s). Vehicle on the right lane still exists (TTC_front_right$\leq$ 6.84 s, TTC_behind_right$\leq$ 3.8 s).

It could be obtained from the description of the critical situations that the reason for the HPRS fluctuation in this period of time is the lane changing maneuver with high speed and the steering wheel frequently turning left and right which indicates that the driver's ability to control these situations is erratic. The driver detected the front vehicle and decided to change to left lane. Therefore, the driver braked to try to decrease the speed, but the braking time was too short and the brake was not applied with sufficient force, the ego-vehicle speed was not actually reduced. The driver completed the lane changing at high vehicle speed environment without enough experience and skills. With the driver error taxonomy in [SS09], the critical behavior of the driver is identified based on Swiss chess model, as shown in Table 5.1

Table 5.1: Analysis of human driver critical behaviors with Swiss chess model [He22a]

Critical behaviors	Active behaviors		Latent conditions
	External error mode	Underlying psychological mechanism	
Fail to press the brake enough	Action too little	Action execution of action error	Time pressure, unfamiliar with driving simulator (inexperience)
Misjudge ego-vehicle speed	Misjudgment	Situation assessment of cognitive and decision-making error	

The latent conditions of the critical behaviors are identified as time pressure and unfamiliar with driving simulator. The TTC_front is decreasing to the lowest as 1.69 s indicates the driver has less time redundancy to complete actions related to driving safety, such as reading the dashboard to avoid misjudge the ego-vehicle speed. Moreover, although participant_1 is experienced in real driving context, his experience on driving simulator is zero, especially unfamiliar with the steering wheel, gas pedal, and braking force needed to control the simulator. For example, the steering wheel of the simulator is more sensitive than real vehicles, and the braking needs to be pressed harder to slow down the vehicle. This unfamiliar with driving simulator induces the critical behaviors like failing to press the brake enough.

The other direction is slice-hole framework with the slices indicating the CPCs and the holes presenting human drivers' experience and skills on situations. In this case, 10 CPCs generated in the modified fuzzy-based CREAM approach are assigned to

10 slices in the modified Swiss chess model and the holes in each slice are related to the membership functions with different effects on performance reliability. When the holes in the slices momentarily line up to a trajectory to the final slice, a HPRS result is obtained. With the continuously generation of HPRS, the HPRS results with time is achieved. It should be noted that the outcomes (e.g. HPRS) from the modified Swiss chess model are not necessarily critical results since holes that are passed through may also correspond to membership functions with positive or not significant effects. Only when more holes with negative effects on performance reliability are passed through, the final outcomes (e.g. HPRS) are indicating critical situations which could be evaluated by control modes in CREAM approach. Therefore, maintaining safe driving behaviors mean keeping driving behavior trajectory through as many holes with positive or not significant effects as possible. Meanwhile, driving behaviors always pass through the holes in slides, so the final HPRS results could be obtained. The traversal of the holes to obtain the final HPRS explains why the HPRS is the summation of different CPC scores.

5.7 Example: HPRS for situated and personalized monitoring of human behaviors

With the increased proportion of human-related accidents in industry and traffic fields, the interest in using assistance systems for supervision of human operators is increasing [SHS18]. Supervision of human behaviors often focuses on the detection of operating errors, unauthorized actions, or implicitly on the violation of protection goals [FDJR19]. Many assistance systems are developed to monitor human operator behaviors and states in different application fields. In [FTAW20], an architecture for human supervision of automation in aviation is proposed which includes the actions of both a human pilot and an autopilot to ensure resilient tracking performance when anomalies occur. In maritime surveillance, a user study conceptualizing knowledge is implemented to support operators' situation awareness for enabling the possibility to detect anomalous behaviors [NVLZE08].

The architecture of situation-operation-modeling (SOM) for interaction of intelligent and autonomous systems is developed to realize the automated supervision of human-machine-interaction [AS08]. In the past this approach has been applied to dynamic driving context. In [FS12], the lane changing maneuver is supervised with SOM approach by interpreting the driving scene and driver action with 'situation' and 'operator'. The main result of this paper is defining individualizable criteria for the decision moment when individuals as deciding to pass (start overtaking maneuver), so initializing a new action changing the upcoming action options. In [SHS18], a fuzzy SOM approach is developed for modeling interaction-based knowledge structures to handle event-discrete situations in a simulated driving environment and to automatically generate a full and individualized knowledge space of sets of situations and actions and related individualized conditions. Using the SOM approach, action space could be generated with possible actions the operator could make considering available options [EGVS10]. The research gap in the existing SOM-based monitoring approach is to automatically integrate individualized criteria for the evaluation of action sequences into the situationally generated action space. In this way, it would be possible to automatically evaluate whether specific action sequences are safe or rather unsafe for this person, e.g., because the action sequence is particularly familiar to this person or because actions/constellations foreseeably occur in the intended action or in the action space that are unsafe or with which the person is not familiar or which he or she demonstrably cannot master. Such an additional option would improve assistance in human-machine interaction and lead to more reliable human-machine systems.

5.7.1 SOM-based human performance reliability evaluation

Situation-operator-modeling

A situation-operator-modeling approach is developed in [Söf01a] allowing the modeling of human-machine-interaction and to map the changes and scenes from the real world to a graph-based-model. Changes are modeled as sequences consisting of items scenes and items actions. A scene is modeled as a situation and an action as an operator. In Figure 5.23 a SOM-based sequence is shown consisting of an actual situation S_i, a current operator O_i and the following situation S_{i+1}. An operator is represented as a white ellipse. A situation is described as a situation vector represented as gray ellipse.

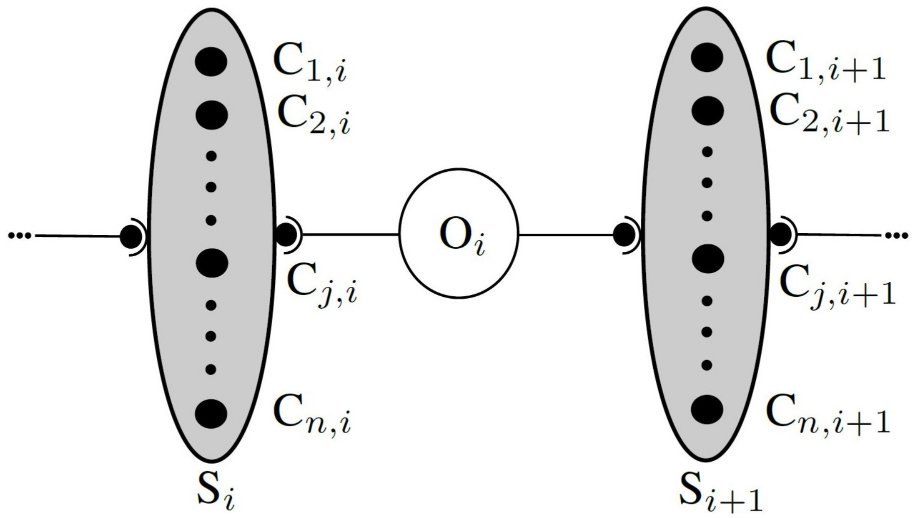

Figure 5.23: Situation-operator-situation sequence [Söf01a]

Table 5.2: List of characteristics including in the situation vector [HBS22]

Characteristic	Unit
C_1 : Longitudinal speed	[km/h]
C_2 : Lateral speed	[km/h]
C_3 : Longitudinal acceleration	[m/s^2]
C_4 : Lateral acceleration	[m/s^2]
C_5 : Yaw angle	[°]
C_6 : Steering wheel angle	[°]
C_7 : Direction indicator to the left	[-]
C_8 : Direction indicator to the right	[-]
C_9 : Lane number	[-]
C_{10}: TTC to front vehicle	[s]
C_{11}: Driving area in the left lane	[-]
C_{12}: Driving area in the right lane	[-]
C_{13}: Distance to front vehicle	[m]

A situation S_i includes a set of characteristics $C_{j,i}$, can be physical, logical, functional, or informational terms and is expressed by its related values. A situation is related to a fix problem configuration.

Using the SOM-approach actions in the real world are modeled as operators. An operator is related to its functionality F, which depends of explicit and implicit assumptions. The assumptions are described by suitable mathematical, logical, or textual expressions. A current situation S_i and the following situation S_{i+1} are connected by an operator, so that an operator can effect the structure and the values related to the characteristics in the following situation.

Operators and characteristics

In the Table 5.2 the characteristics included in a situation vector are shown.

The characteristics C_7 and C_8 provide the statement about the direction indicator and have a Boolean type. If the direction indicator to any direction (left or right) is on, the value of the related characteristic (C_7 or C_8) changes from 'False' to 'True'. The characteristic C_9 gives the number, in which lane the ego-vehicle is driving in the current moment. The characteristics C_{11} and C_{12} provide a statement about the availability of the driving area in the left and right lanes close to the ego-vehicle, which are Boolean.

A sequence consisting of items operators and situations, which describes a sequence of actions, can be replaced as a meta-operator. An example of a meta-operator is 'changing to the left lane' shown in the Fig. 5.24. This meta-operator consists of

Table 5.3: List of characteristics of the situation vector [HBS22]

Operator	Description
O_1	Acceleration
O_2	Deceleration
O_3	Keeping the actual speed
O_4	Turn on the left direction indicator
O_5	Turn off left direction indicator
O_6	Turn on right direction indicator
O_7	Turn off the right direction indicator
O_8	Steering to the left
O_9	Steering to the right

the basic operators 'Turn on the left direction indicator' O_4, 'Operate steering wheel to the left' O_8 ,'Turn off left direction indicator' O_5, 'Steering to the right' O_9 ,and 'Turn off left direction indicator' O_5 (cf. Table 5.3), so describes the 'Changing to left lane' sequence (cf. Figure 5.24).

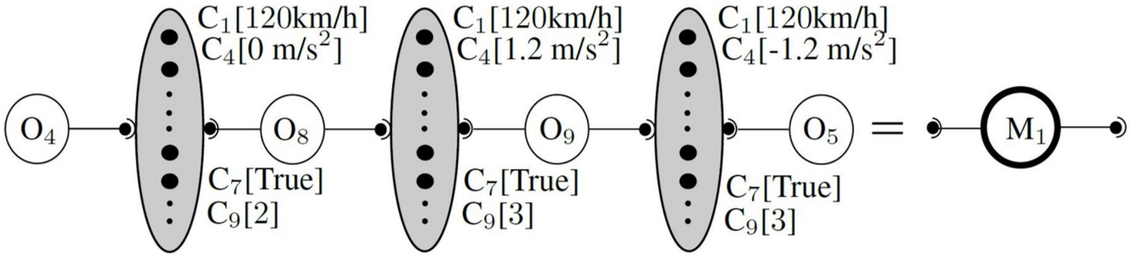

Figure 5.24: Meta-operator 'Changing to the left lane' [HBS22]

The operators describing the actions of the driver are shown in the Table 5.3.

In this example an overtaking maneuver is considered (cf. Figure 5.25). The ego-vehicle is the red vehicle and the vehicle, which has to be overtaken, is the blue vehicle. Possible vehicles driving in the left lane are represented with white color. The final desired situation is, that the ego-vehicle overtakes the blue vehicle considering the environment, and so other vehicles. More than one possibility lead to the final desired situation. Using the situation-operator-modeling an action space consisting of possible driver's behaviors allowing to reach the final desired situation of overtaking the blue vehicle is proposed in the following part.

Action space

For every action taking or decision making moment in time a SOM-based action can be established to map the individual and situated action options for this moment in

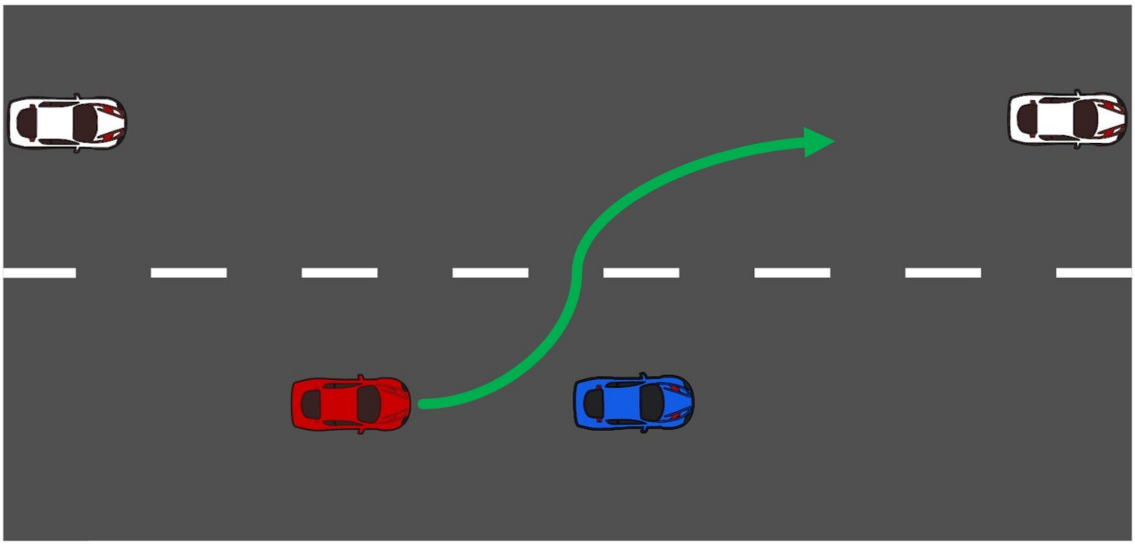

Figure 5.25: Overtaking maneuver on a highway (2 lanes for one direction): Ego-vehicle (red) [HBS22]

time. The intended safety evaluation is based on this continuously changing discrete events. Beside the continuous evaluation of realized actions also the options in specific moments in time can be evaluated establishing a new safety-related performance measure or decision making.

A SOM-based action space consisting of permissible operator sequences, which lead to the desired final situation of overtaking the blue vehicle, is developed and shown in Figure 5.26. In the concrete example, four possible paths lead to the desired final situation and are explained as follows:

Path I: In this case the driver keeps the current speed and waits of the passing of vehicle(s) in the left lane (C_{11} = 'False'). After the left lane is free, the driver changes to the left lane (cf. meta-operator 'Changing to the left lane' in Figure 5.24). The vehicle, which has to be overtaken, accelerates (C_{12} = 'False'), so that the driver of the ego-vehicle has to decelerate and then to keep the current speed. After the vehicle in the right lane do not accelerate and vehicles in the front keep the speed, the driver of the ego-vehicle can overtake by accelerating.

Path II: In this case the driver keeps the current speed and waits of the passing of vehicle(s) in the left lane (C_{11} = 'False'). After the left lane is free, the driver changes to the left lane (cf. meta-operator 'Changing to the left lane' in Figure 5.24). The vehicle in the right lane do not accelerate and vehicles in the front keep the speed, the driver of the ego-vehicle can overtake by accelerating.

Path III: This case is the optimal driving behavior to reach the final desired situation of overtaking the blue vehicle. The left lane is free (C_{12} = 'True'), the driver changes to the left lane (cf. meta-operator 'Changing to the left lane' in Figure

5.24). The vehicle in the right lane do not accelerate and vehicles in the front keep the speed, the driver of the ego-vehicle can overtake by accelerating.

Path IV: The left lane is free (C_{12} = 'True'), the driver changes to the left lane (cf. meta-operator 'Changing to the left lane' in Figure 5.24). The vehicle, which has to be overtaken, accelerates (C_{12} = 'False'), so the driver of the ego-Vehicle has to decelerate and then to keep the current speed. After the vehicle in the right lane do not accelerate and vehicles in the front keep the speed, driver of the ego-vehicle can overtake by accelerating.

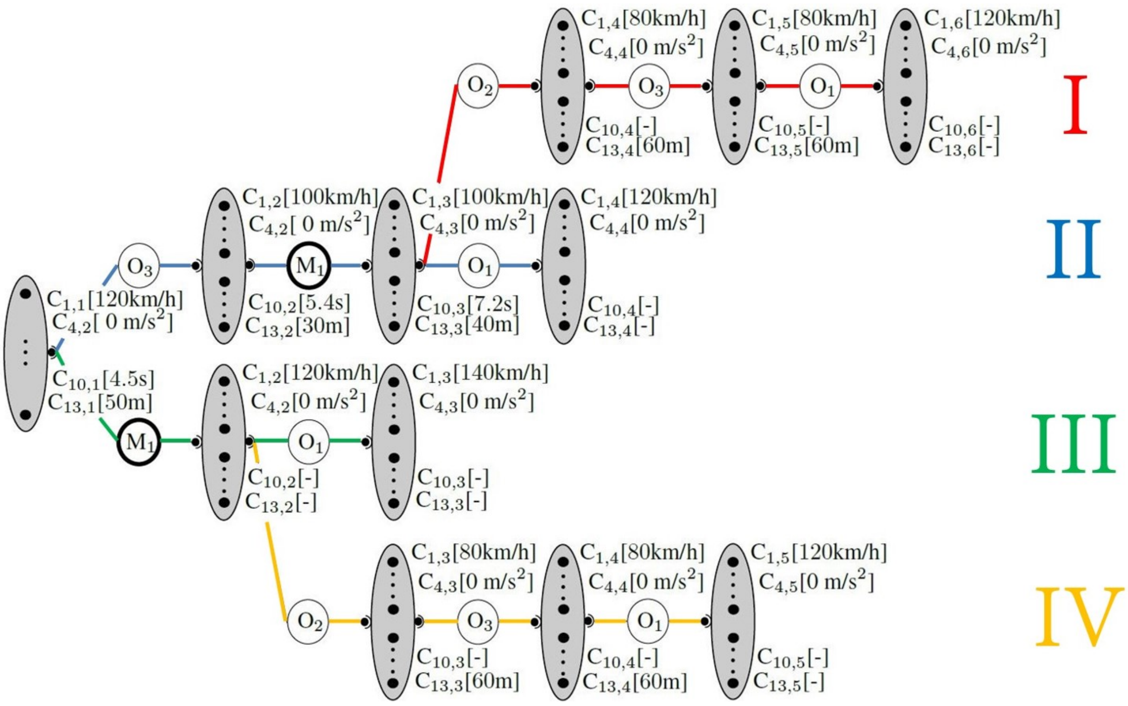

Figure 5.26: SOM-based action space for overtaking [HBS22]

Evaluation options by summarizing safety-related performance scores

To quantitatively evaluate different options and define the optimal action sequence in action space, the group of artificial values of characteristics for the situations in Figure 5.24 and 5.26 are defined. It is assumed that the vehicle speed in front (vehicles in blue and white in front) maintains the fixed speed of 80 km/h, the ego-vehicle speed is varying between 80 and 140 km/h. The distance between the front vehicles and ego-vehicle is from 30 to 60 m considering the traffic rules. In this case, the TTC of front vehicle and ego-vehicle is defined. When the time for lane changing and acceleration is artificially defined as 10 s, the lateral and longitudinal accelerations could be calculated with the relationship of speed and time. With

Table 5.4: HPRS of situations in meta-operator [HBS22]

Operators	HPRS
O_4	3.83
O_8	3.42
O_9	3.42

Table 5.5: HPRS of situations in action space [HBS22]

Path	Operators	HPRS
Path I	O_3	2.88
	M_1	3.31
	O_2	3.43
	O_3	3.63
	O_1	2.63
Path II	O_3	2.88
	M_1	3.31
	O_1	2.83
Path III	M_1	3.00
	O_1	2.83
Path IV	M_1	3.00
	O_2	3.32
	O_3	3.63
	O_1	2.63

the artificial defined values of characteristics in action space for the driving task of overtaking maneuver described in Figure 5.24 and 5.26, the HPRS of each situation could be calculated with the membership functions generated from real driving data. The results is presented in Table 5.4 and Table 5.5.

Different operators result in the changes of CPCs values in the modified CREAM, leading to the HPRS values fluctuation. From Table 5.4, it can be detected that HPRS is decreasing during lane change maneuver as the lateral acceleration is fluctuating. It can be observed in Table 5.5 that path III is the optimal action sequence as it has less action sequences which indicating less cognition requirement and the values of action related HPRS (O_1) is larger than the same action in path I and path IV. Path I dominates most action sequences presenting human driver has more information processing and action implementation, and the action HPRS (O_1) is less than the same action in path II and path III.

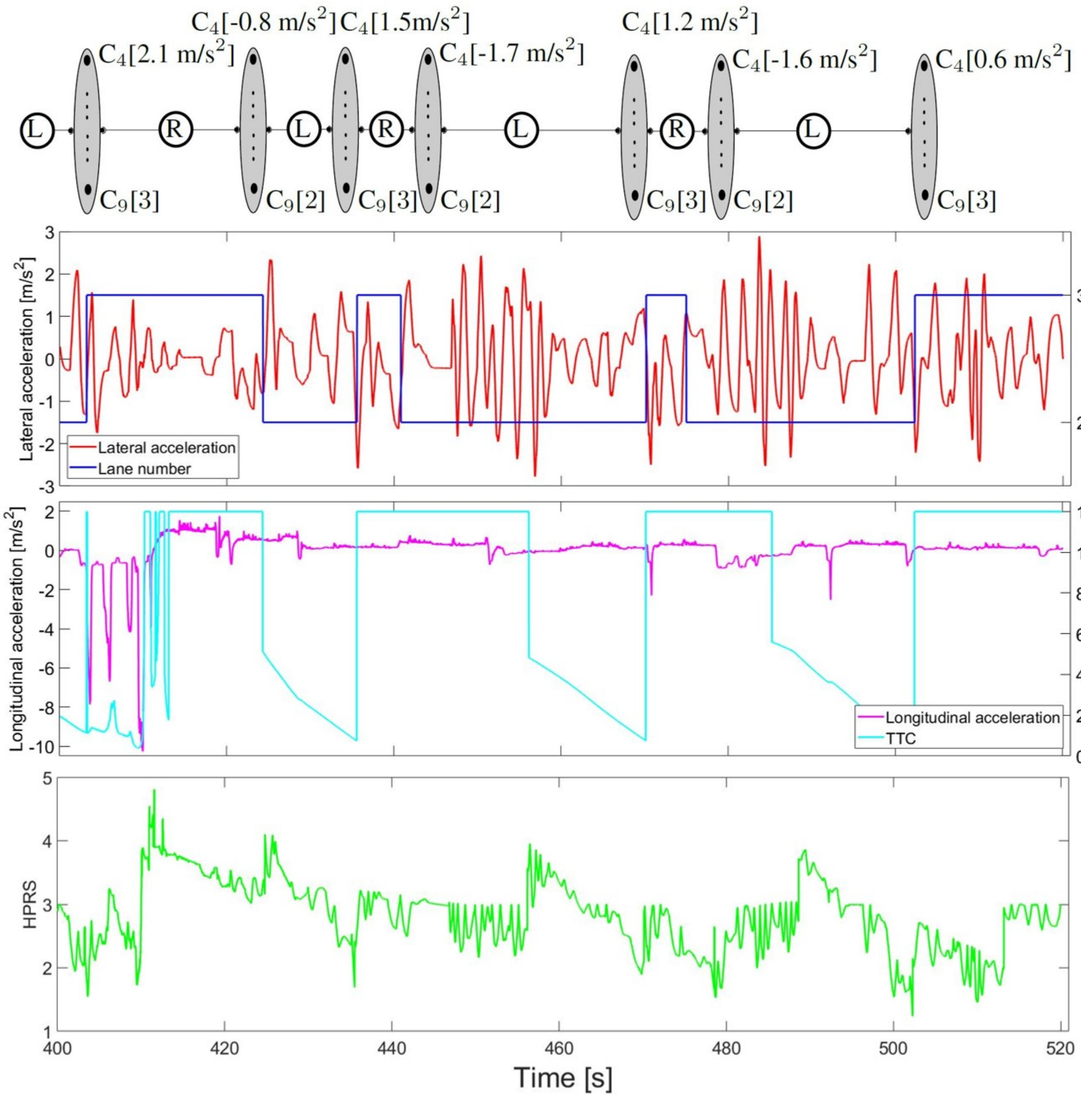

Figure 5.27: Synchronization of SOM-based action sequence and HPRS in lane changing maneuver [HBS22]

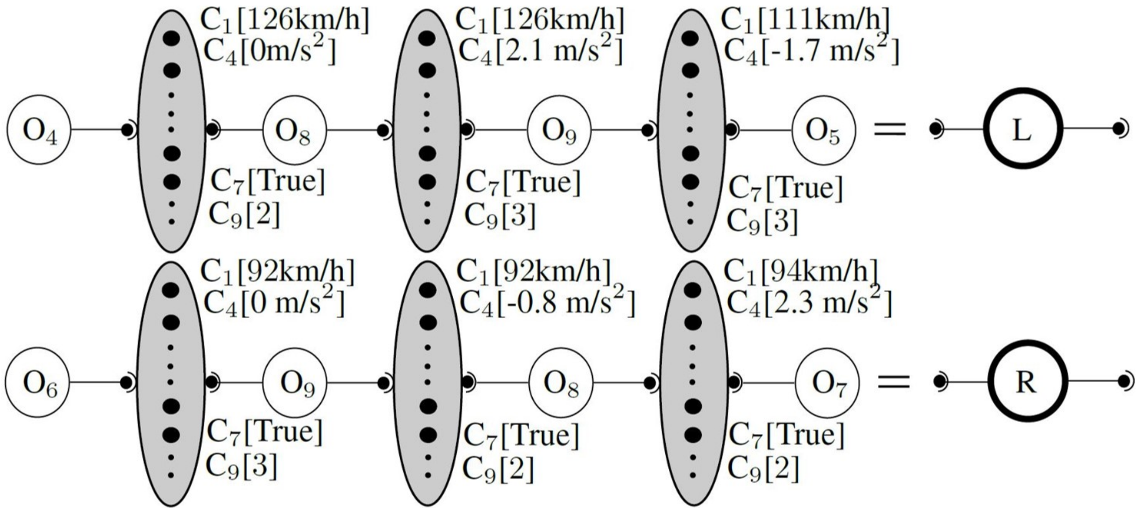

Figure 5.28: Meta-operator of lane changing to left and right in simulated driving [HBS22]

5.7.2 Real-time applicable SOM-based HPRS for real time driver safety evaluation

The SOM approach applying to real driving data in lane changing maneuver is considered. Totally seven lane changing maneuvers including changing to left lane and changing to right lane in the selected data are considered. The changes of lateral acceleration can indicate the time of lane changing as the lateral acceleration is maintaining around 0 m/s^2 when it is lane keeping. The lane changing is a continuous process which start when the lateral acceleration begins to change and ends with the lateral acceleration returning to around 0 m/s^2 and in between the lane number is changed. In this thesis, the top and bottom points in lateral acceleration near the lane number changing are selected as the time of starting and ending of the lane changing maneuver to present the HPRS varying. In this case, the synchronization map of SOM based action sequence and HPRS in lane changing maneuver is presented in Figure 5.27.

In Figure 5.27, the HPRS in the selected time synchronizing with the action sequence of lane changing is presented. With the meta-operator of lane changing to left and lane changing to right and the related situations as shown in Figure 5.28, the action sequence of lane changing is illustrated. Meanwhile, the corresponding HPRS during the lane changing period is presented synchronously. In this case, the driver's lane changing behavior is monitored and evaluated with the SOM-based HPRS approach in real time.

5.8 Summary

In this chapter, the driving data are collected from professional driving simulator. Three data clustering approaches including FN-DBSCAN, CLUSTERDB*, and GMFPE are applied to driving data to automatically determine the membership functions which are assigned to different effects (improved, not significant, and reduced) on performance reliability, so the CPC scores of each CPC is determined and the final HPRS results are calculated with the sum of all CPC score. The HPRS results from three data clustering approach are compared and discussed. It is concluded that the HPRS results with FN-DBSCAN and CLUSTERDB* could more accurately characterize the individualized driving performance as driving data are clustered based on the characteristics of the data themselves and the number of membership functions from GMFPE needs to be predefined. Therefore, the data-driven-based human reliability analysis for human driver performance is realized. The features of the new and dynamic approach are summarized. Swiss chess model could be modified with two directions for the analysis of human driver critical behaviors in dynamic changing situations. One is the typical active behavior and latent conditions structure. The frequently fluctuating HPRS results are selected for the active-latent structure analysis. The critical behaviors are identified with detailed explanation of critical behaviors. In this case, the active behaviors and latent conditions are defined. The other direction only extracts the slices-hole framework from Swiss chess model with defining slices as CPCs and holes as different levels of experience and skills drivers occupied. When holes in the slices are momentarily lining up to a trajectory, the HPRS results are obtained. This framework explains the philosophy that HPRS is the summation of different CPC scores. The example of SOM-based HPRS for situated and personalized monitoring of human driver behaviors are explained. The HEP intervals transition between CREAM approach and SRK model is established and a new SRK related HEP intervals for the evaluation of HPRS results are determined. The HPRS results are evaluated with SRK behavior levels.

6 Summary, conclusion, and outlook

6.1 Summary

Humans are always somewhere integrated in the loops although the continuously development of automation levels in human-machine systems. Human performance reliability is crucial for human-machine systems especially for safety-relevant tasks and operations. In this thesis, the main contents could be summarized as following points.

- A modified fuzzy-based CREAM approach is established for quantitatively calculation of human performance reliability. A new concept to evaluate human performance reliability defined as human performance reliability score (HPRS) is proposed for the quantitative and dynamic evaluation of individualized human performance with time in situated context.
- The situated driving context is taken as an example for the application of the new proposed approach. A new common performance conditions (CPCs) list is generated to describe the main features of situated driving context. With three data clustering approaches (FN-DBSCAN, CLUSTERDB*, and GMFPE), the driving data are clustered to define the parameters of membership functions. The HPRS is finally calculated with the sum of all CPC scores. The HPRS results from three clustering approaches are compared.
- Human error probability (HEP) from three databases (THERP, SRS-HRA, and NARA) are collected to quantify human behavior levels in SRK model. A graphical framework presenting the relationship between human behaviors and HEP is generated. With the analysis of effects of time pressure and training on SRK levels switching, a more general map describing SRK levels switching is proposed and the possible applications are discussed. The detection and evaluation of critical human driver performance are realized with the quantified SRK levels.
- Taken as an example for the monitoring of situated and personalized human performance, the situation-operator-modeling (SOM) combing with HPRS is presented. With different action options to reach the goal, the action space is generated and the optimal option is determined with HPRS.

6.2 Conclusion

In this thesis, a modified fuzzy-based CREAM appraoch is established for the evaluation of human performance reliability in dynamic changing situations. To determine the critical performance in situated driving, human behavior levels in SRK model are defined. The HEP intervals of each behavior level (skill-, rule-, knowledge-based) are determined with HRA databases. The main work is concluded as follows:

- The development of HRA methods with 'three generations' are explained. Three mainly gaps in the existing HRA methods are discussed, including the lack of data for model development and validation, missing consideration of dynamic human reliability in situated context, and heavily reliance on expert knowledge in human reliability evaluation process.

- The detailed steps to establish the modified fuzzy-based CREAM approach for evaluation of dynamic changing situations include: i) the definition of CPCs to describe the main factors of the context; ii) data clustering approaches are executed to generate membership functions; iii) membership functions are assigned to different CPC levels to calculate CPC scores; iv) all ten CPCs scores are added up to obtain the final HPRS. The static HRA and dynamic HRA are discussed to explain that a new and dynamic approach needs to be established because the event evolution is building dynamically as a result of ongoing action and time is considered in dynamic HRA. In this case, situated driving context is selected as the example to indicate the dynamic context as driving behavior data are easily collected in driving simulator. The new list of CPCs describing the main features of situated driving context is defined, which contains ego-vehicle states (longitudinal speed, lateral speed, longitudinal acceleration, and lateral acceleration) and surrounding environment states (TTC front, TTC front left, TTC front right, TTC behind, TTC behind left, and TTC behind right).

- The driving data from two participants are analyzed and the membership functions, CPC scores, and HPRS results with three clustering approaches are obtained. The HPRS results with FN-DBSCAN and CLUSTERDB* could more accurately characterize the individualized driving performance as driving data are clustered based on the characteristics of the data themselves and the number of membership functions from GMFPE needs to be predefined. Therefore, the data driven-based human reliability analysis for human driver performance is realized.

- To detect the critical behavior, the HEP intervals related to skill-, rule-, and knowledge-based levels in SRK model are quantified with three HRA databases (THERP, SRS-HRA, and NARA), as shown in Table 3.7. It could be obtained that the behavior levels are overlapping which denotes the interaction between levels. A graphical framework presenting the relationship between human behaviors and HEP is generated, as presented in Figure 3.6. With the analysis of the effects of two PSFs, time pressure and training, on levels switching, six directions are concluded to indicate the relationship between HEP and typical human behaviors. Direction I/II mean that the quality of the tasks is different but HEP is identical. Directions III/IV indicate that the quality of tasks is identical, while human operators' experience level is varying. Directions V/VI present loosing experience (V) and typical learning process (VI). Based the quantified SRK levels, the CREAM related HPRS is transformed into the SRK related HPRS, which is shown in Figure 5.20. In this case, the critical performance in situated driving is detected.

- Swiss Cheese model could be modified with two directions for the analysis of human driver critical behaviors in dynamic changing situations. One is the typical active behavior and latent conditions structure. The frequently fluctuating HPRS results are selected for the active-latent structure analysis. The critical behaviors are identified with detailed explanation of critical behaviors. In this case, the active behaviors and latent conditions are defined. The other direction only extracts the slices-hole framework from Swiss Cheese model with defining slices as CPCs and holes as different levels of experience and skills drivers occupied. When holes in the slices are momentarily lining up to a trajectory, the HPRS results are obtained. This framework explains the philosophy that HPRS is the summation of different CPC scores.
- As an example to explain the situated and personalized monitoring of human performance with HPRS, a SOM-based human reliability evaluation approach is presented. An action space of overtaking maneuver is generated to describe different possible action sequences and options human driver available, as shown in Figure 5.26. With the calculation of HPRS of each path, it is obtained that option III denotes the optimal action sequence as it has less action sequences which indicating less cognition requirement and the values of situation related HPRS is lower than other paths. The SOM-based action sequence of lane changing maneuver is presented with HPRS synchronously to realize the evaluation of human driver's lane changing performance in real time, which is presented in Figure 5.27.

6.3 Outlook

In the next steps, with the SRK level switching map, a human performance reliability monitoring system can be established combing with modified fuzzy-based CREAM approach, and a individual recognition and evaluation system of training status can be generated with collected operator training data.

Moreover, larger timescale of driving data from human drivers could be adopted to define the membership functions of CPCs to better characterize human driving behaviors. For principal demonstration in this thesis, 120 s of driving data are clustered and analyzed, so the membership functions of CPCs are defined by the driving features of 120 s of driving data, driving behavior characteristics that do not fall within this data range are not discussed. With more data available, more comprehensive driving behavior could be characterized in membership functions. As a result, HPRS results that better indicate human driver performance reliability can be calculated. Moreover, this new established approach could be applied in other human-in-loop fileds, such as the captain performance reliability (sailing task) evaluation.

Bibliography

[AG03] Ackerman, D. S. ; Gross, B. L.: Is time pressure all bad? Measuring the relationship between free time availability and student performance and perceptions. In: *Marketing Education Review* 13 (2003), no. 2, pp. 21–32

[AS08] Ahle, E. ; Söffker, D.: Interaction of intelligent and autonomous systems–part II: realization of cognitive technical systems. In: *Mathematical and Computer Modelling of Dynamical Systems* 14 (2008), no. 4, pp. 319–339

[Bai83] Bainbridge, L.: Ironies of automation. In: *Analysis, design and evaluation of man–machine systems*. Elsevier, 1983, pp. 129–135

[BB08] Balas, V. E. ; Balas, M. M.: Constant time to collision platoons. In: *Int. J. of Computers, Communications and Control* 3 (2008), pp. 33–39

[BC10] Bent, J. ; Chan, K.: Flight training and simulation as safety generators. In: *Human factors in aviation*. Elsevier, 2010, pp. 293–334

[BGJ07] Boring, R. L. ; Griffith, C. D. ; Joe, J. C.: The measure of human error: Direct and indirect performance shaping factors. In: *2007 IEEE 8th Human Factors and Power Plants and HPRCT 13th Annual Meeting* IEEE, 2007, pp. 170–176

[BH09] Bell, J. ; Holroyd, J.: Review of human reliability assessment methods. In: *Health & Safety Laboratory* 78 (2009)

[BHO+94] Benhardt, H. C. ; Held, J. E. ; Olsen, L. M. ; Vail, R. E. ; Eide, S. A.: Savannah river site human error data base development for nonreactor nuclear facilities / Westinghouse Savannah River Co., Aiken, SC (United States). 1994. – Technical Report

[Bor07] Boring, R. L.: Dynamic human reliability analysis: Benefits and challenges of simulating human performance / Idaho National Lab.(INL), Idaho Falls, ID (United States). 2007. – Technical Report

[Bor10] Boring, R. L.: How many performance shaping factors are necessary for human reliability analysis? / Idaho National Lab.(INL), Idaho Falls, ID (United States). 2010. – Technical Report

[Bor12] Boring, R. L.: Fifty years of THERP and human reliability analysis / Idaho National Lab.(INL), Idaho Falls, ID (United States). 2012. – Technical Report

[BR11] Bella, F. ; Russo, R.: A collision warning system for rear-end collision: a driving simulator study. In: *Procedia-social and behavioral sciences* 20 (2011), pp. 676–686

[BR16] Boring, R. L. ; Rasmussen, M.: GOMS-HRA: A method for treating subtasks in dynamic human reliability analysis. In: *Proceedings of the 2016 European Safety and Reliability Conference*, 2016, pp. 956–963

[CCM+06] Chandler, F. T. ; Chang, Y. H. J. ; Mosleh, A. ; Marble, J. L. ; Boring, R. L. ; Gertman, D.: Human reliability analysis methods: selection guidance for NASA. In: *NASA Office of Safety and Mission Assurance, Washington, DC* 123 (2006)

[CDD+92] Cacciabue, P. C. ; Decortis, F. ; Drozdowicz, B. ; Masson, M. ; Nordvik, J. P.: COSIMO: A cognitive simulation model of human decision making and behavior in accident management of complex plants. In: *IEEE Transactions on Systems, Man, and Cybernetics* 22 (1992), no. 5, pp. 1058–1074

[CGW13] Casner, S. M. ; Geven, R. W. ; Williams, K. T.: The effectiveness of airline pilot training for abnormal events. In: *Human factors* 55 (2013), no. 3, pp. 477–485

[CH19] Casner, S. M. ; Hutchins, E. L.: What do we tell the drivers? Toward minimum driver training standards for partially automated cars. In: *Journal of cognitive engineering and decision making* 13 (2019), no. 2, pp. 55–66

[CRSWP96] Cooper, S. E. ; Ramey-Smith, A. M. ; Wreathall, J. ; Parry, G. W.: A technique for human error analysis (ATHEANA) / Nuclear Regulatory Commission. 1996. – Technical Report

[CSM15] Christensen, W. ; Sutton, J. ; McIlwain, D.: Putting pressure on theories of choking: Towards an expanded perspective on breakdown in skilled performance. In: *Phenomenology and the Cognitive Sciences* 14 (2015), no. 2, pp. 253–293

[DBD17] De Boer, R. ; Dekker, S.: Models of automation surprise: results of a field survey in aviation. In: *Safety* 3 (2017), no. 3, pp. 20

[DC96] Dougherty, E. M. ; Collins, E. P.: Assessing the reliability of skilled performance. In: *Reliability Engineering & System Safety* 51 (1996), no. 1, pp. 35–42

[DDWN09] Dahlstrom, N. ; Dekker, S. ; Winsen, R. van ; Nyce, J.: Fidelity and validity of simulator training. In: *Theoretical Issues in Ergonomics Science* 10 (2009), no. 4, pp. 305–314

[Dek17] Dekker, S.: *The field guide to understanding 'human error'.* CRC press, 2017

[Dhi07] Dhillon, B. S.: *Human reliability and error in transportation systems.* Springer Science & Business Media, 2007

[Dhi17] Dhillon, B. S.: *Safety, reliability, human factors, and human error in nuclear power plants.* CRC Press Florida, 2017

[DHO08] Derbel, I. ; Hachani, N. ; Ounelli, H.: Membership functions generation based on density function. In: *2008 International Conference on Computational Intelligence and Security* vol. 1 IEEE, 2008, pp. 96–101

[DHUM10] Dong, Y. ; Hu, Z. ; Uchimura, K. ; Murayama, N.: Driver inattention monitoring system for intelligent vehicles: A review. In: *IEEE transactions on intelligent transportation systems* 12 (2010), no. 2, pp. 596–614

[Dia11] Diaz, A. B.: The training of operating personnel at Spanish nuclear power plant. In: *2011 International Nuclear Atlantic Conference - INAC 2011* (2011), pp. 1–8

[DPIMR13] Di Pasquale, V. ; Iannone, R. ; Miranda, S. ; Riemma, S.: An overview of human reliability analysis techniques in manufacturing operations. In: *Operations management* (2013), pp. 221–240

[DPMIR15] Di Pasquale, V. ; Miranda, S. ; Iannone, R. ; Riemma, S.: A simulator for human error probability analysis (SHERPA). In: *Reliability Engineering & System Safety* 139 (2015), pp. 17–32

[E+84] Embrey, D. E. [u. a.]: SLIM-MAUD (Success Likelihood Index Methodology Multi-Attribute Utility Decomposition): An approach to assessing human error probabilities using structured expert judgement. In: *NUREG/CR3518* 2 (1984)

[EGVS10] Ertle, P. ; Gamrad, D. ; Voos, H. ; Söffker, D.: Action planning for autonomous systems with respect to safety aspects. In: *2010 IEEE International Conference on Systems, Man and Cybernetics* IEEE, 2010, pp. 2465–2472

[EMP16] Eboli, L. ; Mazzulla, G. ; Pungillo, G.: Combining speed and acceleration to define car users' safe or unsafe driving behaviour. In: *Transportation research part C: emerging technologies* 68 (2016), pp. 113–125

[End17] Endsley, M. R.: Toward a theory of situation awareness in dynamic systems. In: *Situational awareness*. Routledge, 2017, pp. 9–42

[FDJR19] Fridman, L. ; Ding, L. ; Jenik, B. ; Reimer, B.: Arguing machines: Human supervision of black box AI systems that make life-critical decisions. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0

[FHLR20] Fisher, D. L. ; Horrey, W. J. ; Lee, J. D. ; Regan, M. A.: *Handbook of human factors for automated, connected, and intelligent vehicles*. CRC Press, 2020

[FP16] Fleming, E. ; Pritchett, A.: SRK as a framework for the development of training for effective interaction with multi-level automation. In: *Cognition, Technology & Work* 18 (2016), no. 3, pp. 511–528

[FS12] Fu, X. G. ; Söffker, D.: Modeling of individualized cognitive driving supervision for intelligent vehicles. In: *2012 IEEE Vehicle Power and Propulsion Conference* IEEE, 2012, pp. 1317–1322

[FTAW20] Farjadian, A. B. ; Thomsen, B. ; Annaswamy, A. M. ; Woods, D. D.: Resilient flight control: An architecture for human supervision of automation. In: *IEEE Transactions on Control Systems Technology* 29 (2020), no. 1, pp. 29–42

[FZBD+18] Fan, S. Q. ; Zhang, J. F. ; Blanco-Davis, E. ; Yang, Z. L. ; Wang, J. ; Yan, X. P.: Effects of seafarers' emotion on human performance using bridge simulation. In: *Ocean Engineering* 170 (2018), pp. 111–119

[GB93] Gertman, D. I. ; Blackman, H. S.: *Human reliability and safety analysis data handbook*. John Wiley & Sons, 1993

[GBM+05] Gertman, D. ; Blackman, H. ; Marble, J. ; Byers, J. ; Smith, C. [u. a.]: The SPAR-H human reliability analysis method. In: *US Nuclear Regulatory Commission* 230 (2005), no. 4, pp. 35

[GKK+17] Gibson, W. H. ; Kirwan, B. ; Kennedy, R. ; Edmunds, J. ; Umbers, I.: Nuclear Action Reliability Assessment (NARA), further development of a data-based HRA tool. In: *Contemporary Ergonomics 2008*. Taylor & Francis, 2017, pp. 182–187

[GKLB16] Gold, C. ; Körber, M. ; Lechner, D. ; Bengler, K.: Taking over control from highly automated vehicles in complex traffic situations: the role of traffic density. In: *Human factors* 58 (2016), no. 4, pp. 642–652

[GM11] Griffith, C. D. ; Mahadevan, S.: Inclusion of fatigue effects in human reliability analysis. In: *Reliability Engineering & System Safety* 96 (2011), no. 11, pp. 1437–1447

[Gre86] Grefenstette, J. J.: Optimization of control parameters for genetic algorithms. In: *IEEE Transactions on systems, man, and cybernetics* 16 (1986), no. 1, pp. 122–128

[GSM19] Groth, K. M. ; Smith, R. ; Moradi, R.: A hybrid algorithm for developing third generation HRA methods using simulator data, causal models, and cognitive science. In: *Reliability Engineering & System Safety* 191 (2019), pp. 106507

[Háj13] Hájek, P.: *Metamathematics of fuzzy logic*. vol. 4. Springer Science & Business Media, 2013

[HCS18] Hsiao, H. W. ; Chang, J. ; Simeonov, P.: Preventing emergency vehicle crashes: Status and challenges of human factors issues. In: *Human factors* 60 (2018), no. 7, pp. 1048–1072

[HGW+11] Hooey, B. L. ; Gore, B. F. ; Wickens, C. D. ; Scott-Nash, S. ; Socash, C. ; Salud, E. ; Foyle, D. C.: Modeling pilot situation awareness. In: *Human modelling in assisted transportation*. Springer, 2011, pp. 207–213

[HLL92] Hahn, M. ; Lawson, R. ; Lee, Y. G.: The effects of time pressure and information load on decision quality. In: *Psychology & Marketing* 9 (1992), no. 5, pp. 365–378

[HLLS21] He, C. ; Lum, Y. Y. ; Lee, K. Y. ; Söffker, D.: Human reliability estimation based on fuzzy logic-modified CREAM approach. In: *2021 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)* IEEE, 2021, pp. 45–50

[HMB+09] Hinkelmann, K. ; Moritz, S. ; Botzenhardt, J. ; Riedesel, K. ; Wiedemann, K. ; Kellner, M. ; Otte, C.: Cognitive impairment in major depression: association with salivary cortisol. In: *Biological psychiatry* 66 (2009), no. 9, pp. 879–885

[HO07] Hachani, N. ; Ounelli, H.: Improving cluster method quality by validity indices. In: *Flairs Conference*, 2007, pp. 479–483

[Hol98] Hollnagel, E.: *Cognitive reliability and error analysis method (CREAM)*. Elsevier, 1998

[HSL85] Hannaman, G. W. ; Spurgin, A. J. ; Lukic, Y.: A model for assessing human cognitive reliability in PRA studies. In: *Conference record for 1985 IEEE third conference on human factors and nuclear safety*, 1985, pp. 343–353

[Hwa94] Hwang, M. I.: Decision making under time pressure: A model for information systems research. In: *Information & Management* 27 (1994), no. 4, pp. 197–203

[JPK+20] Jung, W. ; Park, J. ; Kim, Y. ; Choi, S. Y. ; Kim, S.: HuREX–A framework of HRA data collection from simulators in nuclear power plants. In: *Reliability Engineering & System Safety* 194 (2020), pp. 106235

[JSXG10] Jiang, Y. J. ; Sun, Z. Q. ; Xie, H. W. ; Gong, E. L.: A Human Error Probability Quantification Method Based on SRK Framewok. In: *2010 3rd International Conference on Information Management, Innovation Management and Industrial Engineering* vol. 2 IEEE, 2010, pp. 89–92

[KGK+04] Kirwan, B. ; Gibson, H. ; Kennedy, R. ; Edmunds, J. ; Cooksley, G. ; Umbers, I.: Nuclear action reliability assessment (NARA): a data-based HRA tool. In: *Probabilistic safety assessment and management* Springer, 2004, pp. 1206–1211

[Kim01] Kim, I. S.: Human reliability analysis in the man–machine interface design review. In: *Annals of nuclear energy* 28 (2001), no. 11, pp. 1069–1081

[KKTAL97] Kirwan, B. ; Kennedy, R. ; Taylor-Adams, S. ; Lambert, B.: The validation of three Human Reliability Quantification techniques—THERP, HEART and JHEDI: Part II—Results of validation exercise. In: *Applied ergonomics* 28 (1997), no. 1, pp. 17–25

[KM13] Kontogiannis, T. ; Malakis, S.: Strategies in controlling, coordinating and adapting performance in air traffic control: Modelling 'loss of control'events. In: *Cognition, technology & work* 15 (2013), no. 2, pp. 153–169

[KMO15] Kyriakidis, M. ; Majumdar, A. ; Ochieng, W. Y.: Data based framework to identify the most significant performance shaping factors in railway operations. In: *Safety science* 78 (2015), pp. 60–76

[KRM+03] Klein, G. ; Ross, K. G. ; Moon, B. M. ; Klein, D. E. ; Hoffman, R. R. ; Hollnagel, E.: Macrocognition. In: *IEEE intelligent systems* 18 (2003), no. 3, pp. 81–85

[KSH06] Kim, M. C. ; Seong, P. H. ; Hollnagel, E.: A probabilistic approach for determining the control mode in CREAM. In: *Reliability Engineering & System Safety* 91 (2006), no. 2, pp. 191–199

[LC15] Le Coze, J. C.: Reflecting on Jens Rasmussen's legacy: A strong program for a hard problem. In: *Safety science* 71 (2015), pp. 123–141

[LKJ12] Lee, S. J. ; Kim, J. ; Jung, W.: A human reliability evaluation tool for main control rooms in nuclear power plants. In: *Proceedings of NPIC & HMIT* (2012), Mai, pp. 1212–1213

[LLC20] Larouzee, J. ; Le Coze, J. C.: Good and bad reasons: The Swiss cheese model and its critics. In: *Safety science* 126 (2020), pp. 104660

[MC04] Mosleh, A. ; Chang, Y. H.: Model-based human reliability analysis: prospects and requirements. In: *Reliability Engineering & System Safety* 83 (2004), no. 2, pp. 241–253

[McC05] McCall, J.: Genetic algorithms for modelling and optimisation. In: *Journal of computational and Applied Mathematics* 184 (2005), no. 1, pp. 205–222

[ME97] Maule, A. J. ; Edland, A. C.: The effects of time pressure on human judgment and decision making. In: *Decision making: Cognitive models and explanations* (1997), pp. 189–204

[MPN+11] Martinie, C. ; Palanque, P. ; Navarre, D. ; Winckler, M. ; Poupart, E.: Model-based training: An approach supporting operability of critical interactive systems. In: *Proceedings of the 3rd ACM SIGCHI symposium on engineering interactive computing systems*, 2011, pp. 53–62

[MSE+21] McDonnell, A. S. ; Simmons, T. G. ; Erickson, G. G. ; Lohani, M. ; Cooper, J. M. ; Strayer, D. L.: This is your brain on Autopilot: Neural indices of driver workload and engagement during partial vehicle automation. In: *Human factors* (2021), pp. 00187208211039091

[MSJ+15] Miller, D. ; Sun, A. ; Johns, M. ; Ive, H. ; Sirkin, D. ; Aich, S. ; Ju, W.: Distraction becomes engagement in automated driving. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* vol. 59 Sage Publications Sage CA: Los Angeles, CA, 2015, pp. 1676–1680

[NNAV21] Nidamanuri, J. ; Nibhanupudi, C. ; Assfalg, R. ; Venkataraman, H.: A Progressive Review-Emerging Technologies for ADAS Driven Solutions. In: *IEEE Transactions on Intelligent Vehicles* (2021), pp. 1–18

[NU09] Nasibov, E. N. ; Ulutagay, G.: Robustness of density-based clustering methods with various neighborhood relations. In: *Fuzzy Sets and Systems* 160 (2009), no. 24, pp. 3601–3615

[NVLZE08] Nilsson, M. ; Van Laere, J. ; Ziemke, T. ; Edlund, J.: Extracting rules from expert operators to support situation awareness in maritime surveillance. In: *2008 11th International conference on information fusion* IEEE, 2008, pp. 1–8

[Phi18] Philippart, M.: Human reliability analysis methods and tools. In: *Space Safety and Human Performance.* Elsevier, 2018, pp. 501–568

[PJK20] Park, J. ; Jung, W. ; Kim, J.: Inter-relationships between performance shaping factors for human reliability analysis of nuclear power plants. In: *Nuclear Engineering and Technology* 52 (2020), no. 1, pp. 87–100

[PLH17] Pan, X. ; Lin, Y. ; He, C. J.: A review of cognitive models in human reliability analysis. In: *Quality and Reliability Engineering International* 33 (2017), no. 7, pp. 1299–1316

[PMS15] Pinto, J. M. de O. ; Melo, P. F. ; Saldanha, P. L. C.: Human-Machine interface (HMI) scenario quantification performed by ATHEANA, A Technique for Human Error Analysis. In: *Safety and Reliability of Complex Engineered Systems* (2015), pp. 3111–3118

[PP06] Peters, G. A. ; Peters, B. J.: *Human error: Causes and control.* CRC press, 2006

[PSW00] Parasuraman, R. ; Sheridan, T. B. ; Wickens, C. D.: A model for types and levels of human interaction with automation. In: *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans* 30 (2000), no. 3, pp. 286–297

[Ras82] Rasmussen, J.: Human errors: A taxonomy for describing human malfunction in industrial installations. In: *Journal of occupational accidents* 4 (1982), no. 2-4, pp. 311–333

[Ras83] Rasmussen, J.: Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. In: *IEEE transactions on systems, man, and cybernetics* (1983), no. 3, pp. 257–266

[Ras87a] Rasmussen, J.: *Information processing and human-machine interaction. An approach to cognitive engineering.* North-Holland, 1987

[Ras87b] Rasmussen, J.: *Mental models and the control of actions in complex environments.* Risø National Laboratory, 1987

[Rea90] Reason, J.: *Human error.* Cambridge university press, 1990

[Rea00] Reason, J.: Human error: models and management. In: *Bmj* 320 (2000), no. 7237, pp. 768–770

[RH17] Reason, J. ; Hobbs, A.: *Managing maintenance error: A practical guide.* CRC Press, 2017

[RNR+20] Ramezani, A. ; Nazari, T. ; Rabiee, A. ; Hadad, K. ; Faridafshin, M.: Human error probability quantification for NPP post-accident analysis using Cognitive-Based THERP method. In: *Progress in Nuclear Energy* 123 (2020), pp. 103281

[RSN] Ruiz-Sánchez, T. ; NELSON, P.: Application of the ATHEANA methodology for the HRA of a PSA scenario for a BWR nuclear power plant. In: *10th International Conference on Probabilistic Safety Assessment and Management 2010, PSAM 2010*, pp. 1520–1530

[RV89] Rasmussen, J. ; Vicente, K. J.: Coping with human errors through system design: implications for ecological interface design. In: *international Journal of Man-machine Studies* 31 (1989), no. 5, pp. 517–534

[SB15] Silberman, M. L. ; Biech, E.: *Active training: A handbook of techniques, designs, case examples, and tips.* John Wiley & Sons, 2015

[SDH+17] Shappell, S. ; Detwiler, C. ; Holcomb, K. ; Hackworth, C. ; Boquet, A. ; Wiegmann, D. A.: Human error and commercial aviation accidents: an analysis using the human factors analysis and classification system. In: *Human error in aviation.* Routledge, 2017, pp. 73–88

[SG83] Swain, A. D. ; Guttmann, H. E.: Handbook of human-reliability analysis with emphasis on nuclear power plant applications. Final report / Sandia National Labs. 1983. – Technical Report

[She17] Sheridan, T. B.: Musings on models and the genius of Jens Rasmussen. In: *Applied ergonomics* 59 (2017), pp. 598–601

[She21] Sheridan, T. B.: Human supervisory control of automation. In: *Handbook of Human Factors and Ergonomics* (2021), pp. 736–760

[Shi19] Shi, Z. Z.: *Advanced artificial intelligence.* vol. 4. World Scientific, 2019

[SHS18] Sarkheyli-Hägele, A. ; Söffker, D.: Learning and representation of event-discrete situations for individualized situation recognition using fuzzy Situation-Operator Modeling. In: *Engineering Applications of Artificial Intelligence* 72 (2018), pp. 357–367

[Shu19] Shuttleworth, J.: SAE Standards News: J3016 automated-driving graphic update. In: *SAE International* (2019)

[Sin15] Singh, S.: Critical reasons for crashes investigated in the national motor vehicle crash causation survey. 2015. – Technical Report

[SM93] Svenson, O. ; Maule, A. J.: *Time pressure and stress in human judgment and decision making.* Springer Science & Business Media, 1993

[SMC10] Salas, E. ; Maurino, D. ; Curtis, M.: Human factors in aviation: An overview. In: *Human factors in aviation* (2010), pp. 3–19

[Söf01a] Söffker, D.: From human–machine-interaction modeling to new concepts constructing autonomous systems: A phenomenological engineering-oriented approach. In: *Journal of Intelligent and Robotic Systems* 32 (2001), no. 2, pp. 191–205

[Söf01b] Söffker, D.: *System-theoretic modeling of the human interaction with technical systems*, Habilitation Thesis (in German), University of Wuppertal, Diss., 2001

[SS09] Stanton, N. A. ; Salmon, P. M.: Human error taxonomies applied to driving: A generic driver error taxonomy and its implications for intelligent transport systems. In: *Safety Science* 47 (2009), no. 2, pp. 227–237

[SSBD14] Shalev-Shwartz, S. ; Ben-David, S.: *Understanding machine learning: From theory to algorithms.* Cambridge university press, 2014

[ST92] Sheridan, T. B. ; Telerobotics, Automation: *Human supervisory control.* 1992

[Str19] Sträter, O.: Human Reliability Assessment—State of the Art for Task-and Goal-Related Behavior. In: *Risk Based Technologies.* Springer, 2019, pp. 143–171

[SV78] Sheridan, T. B. ; Verplank, W. L.: Human and computer control of undersea teleoperators / Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab. 1978. – Technical Report

[Swa87] Swain, A. D.: Accident sequence evaluation program: Human reliability analysis procedure / Sandia National Labs., Albuquerque, NM (USA); Nuclear Regulatory Commission. 1987. – Technical Report

[SXSL09] Sun, Z. Q. ; Xie, H. W. ; Shi, X. J. ; Liu, F. Q.: Engineering approach for human error probability quantification. In: *Journal of Systems Engineering and Electronics* 20 (2009), no. 5, pp. 1144–1152

[SY05] Stanton, N. A. ; Young, M. S.: Driver behaviour with adaptive cruise control. In: *Ergonomics* 48 (2005), no. 10, pp. 1294–1313

[TJ85] Trager Jr, T. A.: Case study report on loss of safety system function events. In: *Report No. AEOD C* 504 (1985)

[TS19] Tanshi, F. ; Söffker, D.: Modeling of takeover variables with respect to driver situation awareness and workload for intelligent driver assistance. In: *2019 IEEE Intelligent Vehicles Symposium (IV)* IEEE, 2019, pp. 1667–1672

[UN08] Ulutagay, G. ; Nasibov, E.: FN-DBSCAN: A novel density-based clustering method with fuzzy neighborhood relations. In: *8th International Conference on Application of Fuzzy Systems and Soft Computing (ICAFS-2008)*, 2008, pp. 101–110

[VD20] Volz, K. M. ; Dorneich, M. C.: Evaluation of cognitive skill degradation in flight planning. In: *Journal of Cognitive Engineering and Decision Making* 14 (2020), no. 4, pp. 263–287

[VDHH93] Van Der Horst, R. ; Hogema, J.: Time-to-collision and collision avoidance systems. (1993)

[Vic99] Vicente, K. J.: *Cognitive work analysis: Toward safe, productive, and healthy computer-based work*. CRC press, 1999

[Wha16] Whaley, A. M.: *Cognitive basis for human reliability analysis*. US Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, 2016

[WHHB21] Wickens, C. D. ; Helton, W. S. ; Hollands, J. G. ; Banbury, S.: *Engineering psychology and human performance*. Routledge, 2021

[Whi04] Whittingham, R.: *The blame machine: Why human error causes accidents*. Routledge, 2004

[WHLS14] Winner, H. ; Hakuli, S. ; Lotz, F. ; Singer, C.: *Handbook of driver assistance systems*. Springer International Publishing Amsterdam, The Netherlands:, 2014

[Wil88] Williams, J. C.: A data-based method for assessing and reducing human error to improve operational performance. In: *Conference Record for 1988 IEEE Fourth Conference on Human Factors and Power Plants,* IEEE, 1988, pp. 436–450

[Woo09] Woods, D.: Rasmussen's SRK 30 years later: Is human factors best in 3's? In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* vol. 53 SAGE Publications Sage CA: Los Angeles, CA, 2009, pp. 217–221

[WS18] Wang, J. ; Söffker, D.: Bridging gaps among human, assisted, and automated driving with DVIs: A conceptional experimental study. In: *IEEE Transactions on Intelligent Transportation Systems* 20 (2018), no. 6, pp. 2096–2108

[Zad78] Zadeh, L. A.: Fuzzy sets as a basis for a theory of possibility. In: *Fuzzy sets and systems* 1 (1978), no. 1, pp. 3–28

[ZB13] Zeeman, A. S. ; Booysen, M. J.: 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)* IEEE, 2013, pp. 756–761

[ZSG17] Zwirglmaier, K. ; Straub, D. ; Groth, K. M.: Capturing cognitive causal paths in human reliability analysis with Bayesian network models. In: *Reliability Engineering & System Safety* 158 (2017), pp. 117–129

[ZTC+17] Zhang, C. ; Tang, P. B. ; Cooke, N. ; Buchanan, V. ; Yilmaz, A. ; Germain, S. W. S. ; Boring, R. L. ; Akca-Hobbins, S. ; Gupta, A.: Human-centered automation for resilient nuclear power plant outage control. In: *Automation in Construction* 82 (2017), pp. 179–192

[ZWX+17] Zhou, Q. J. ; Wong, Y. D. ; Xu, H. ; Van Thai, V. ; Loh, H. S. ; Yuen, K. F.: An enhanced CREAM with stakeholder-graded protocols for tanker shipping safety application. In: *Safety science* 95 (2017), pp. 140–147

This thesis is based on the studies in the following publications from the authors:

Journal papers

[He22b] He, C.; Söffker, D.: Quantification of human behavior levels by extending Rasmussen's SRK model and the effects of time pressure and training on the levels switching. Heliyon 9(4), 2023

[He22a] He, C.; Söffker, D.: Data-driven-based human reliability analysis for individualized performance with different data clustering approaches for dynamical changing situations. Heliyon, 2023, submitted

Conference papers

[HS22] He, C.; Söffker, D.: Identification of human driver critical behaviors and related reliability evaluation in real time. European Conference on Safety and Reliability (ESREL), Dublin, Ireland, 2022

[HBS22] He, C.; Bejaoui, A. ; Söffker, D.: Situated and personalized monitoring of human operators during complex situations. European Conference on Safety and Reliability (ESREL), Dublin, Ireland, 2022

[HLLS21] He, C.; Lum, Y.; Lee, K.; Söffker, D.: Human reliability estimation based on fuzzy logic-modified CREAM approach. 2021 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA), Tallinn, Estonia, 2021, pp. 45-50

[HLL21] He, C.; Söffker, D.: Human reliability analysis in situated driving context considering human experience using a fuzzy-based clustering approach. 2nd IEEE International Conference on Human-Machine Systems, Magdeburg, Germany, 2021, pp. 8-10

[HS20] He, C.; Söffker, D.: Establishing a modified CREAM approach for reliability evaluation. 2020 IEEE International Conference on Human-Machine Systems (ICHMS), Rome, Italy, 2020, pp. 1-6

[HTS20] He, C.; Tanshi, F.; Söffker, D.: Human online reliability estimation applied to real driving maneuvers. 2020 IEEE Conference on Cognitive and Computational Aspects of Situation management (CogSIMA), Victoria, BC, Canada, 2020, pp. 149-154

Workshop contributions

[He19] He, C.; Söffker, D.: A modified CREAM approach to situated human driving context. 8. Interdisziplinärer Workshop Kognitive Systeme: Mensch, Teams, Systeme und Automaten, Duisburg, März 26-28, 2019

Publications which are not included in this thesis:

Conference paper

[BHS22] Bejaoui, A.; He, C.; Söffker, D.: Novel model-based decision support system for reliable human machine systems in complex situations. Annual Conference of the PHM Society 2022, Nashville, Tennessee, USA, 2022

In the context of research work at the Chair of Dynamics and Control, the following student theses have been supervised by Chao He and Univ.-Prof. Dr.-Ing. Dirk Söffker. Development steps and results of the research work and the student theses are integrated with each other and hence are also part of this thesis.

[Ng21] Ng, K. S.: Human in automation systems: Cognition and characterization of human behavior levels related to human reliability analysis, Bachelor Thesis, February 2021

[Lum20] Lum, Y. Y.: Implementing and comparison of a new fuzzified human reliability analysis approach using driving simulator data, Bachelor Thesis, December 2020

[Lee20] Lee, K. Y.: Establishing a human reliability measure for driver assistant system in MATLAB, Bachelor Thesis, July 2020

The following student theses have been supervised by Chao He and Univ.-Prof. Dr.-Ing. Dirk Söffker, which are not included in this thesis:

[Kum22] Kumar, G. S.: Human factors and human performance in human-automation systems: The effects of automation on human reliability, Bachelor Thesis, October 2022

DOI: 10.17185/duepublico/78427
URN: urn:nbn:de:hbz:465-20230602-095341-6