Mälardalen University Press Licentiate Theses
No. 293

A FRAMEWORK FOR RISK ASSESSMENT IN AUGMENTED REALITY-EQUIPPED SOCIO-TECHNICAL SYSTEMS

Soheila Sheikh Bahaei

2020

School of Innovation, Design and Engineering

ISBN 978-91-7485-470-1
ISSN 1651-9256
Printed by E-Print AB, Stockholm, Sweden

Abstract

New technologies, such as augmented reality (AR) are used to enhance human capabilities and extend human functioning; nevertheless they may cause distraction and incorrect human functioning. Systems including socio entities (such as humans) and technical entities (such as augmented reality) are called socio-technical systems. In order to assess risk in such systems, considering new dependability threats (i.e., faults, errors, and failures) caused by augmented reality is essential. For example, failure of extended human functions is a new type of dependability threat introduced to the system because of new technologies. In particular, it is required to identify these new dependability threats and analyze entities and system behavior to be able to uncover their potential impact.

This thesis aims at providing a framework for risk assessment in AR-equipped socio-technical systems by identifying and classifying human failures including AR-extended human failures and by identifying and classifying faults leading to human failures including AR-caused faults. Our work also provides modeling capabilities for socio-technical systems, to enable modeling of AR-relevant dependability threats used for extending analysis techniques to address the requirements for AR-equipped socio-technical systems analysis. To achieve this, we propose a human function taxonomy by extracting functions from state-of-the-art human failure taxonomies, organizing and harmonizing them in addition to extending the taxonomy by adding AR-extended functions extracted from experiments and studies on augmented reality. Besides, we propose a taxonomy of faults leading to human failures by extracting faults from state-of-the-art taxonomies, organizing and harmonizing them in addition to extending the taxonomy by adding AR-caused faults extracted from studies and experiments on augmented reality. In the context of socio-technical system modeling, AR-extended human functions and AR-caused faults are transformed into

enhanced modeling elements for both human and organizational entities. SafeConcert, which is a metamodel for modeling socio-technical systems, is used as the basis for extension of socio entities modeling elements. This extended metamodel can then be used to augment the risk analysis techniques used for socio-technical systems analysis. Concerto-FLA, which is a risk analysis technique for analyzing socio-technical systems, is used as the basis for analyzing system behavior. We show the applicability of our modeling extensions on academic examples and we also conduct a case study to evaluate the analysis capabilities of the provided extensions.

Sammanfattning

Nya teknologier, som förstärkt verklighet (AR), används för att förbättra mänskliga förmågor och utöka mänskliga funktioner; men de riskerar samtidigt att orsaka distraktion och felaktiga mänskliga reaktioner. System som inkluderar både socio-enheter (som människor) och tekniska enheter (som förstärkt verklighet) kallas socio-tekniska system. För att bedöma risker i sådana system är det viktigt att ta hänsyn till nya tillförlitlighetshot orsakade av förstärkt verklighet. Felaktiga utökade mänskliga funktioner är till exempel en ny typ av beroendehot som introducerats på grund av nya teknologier. Det är särskilt viktigt att identifiera dessa nya tillförlitlighetshot och analysera enheter och systembeteende för att förstå deras potentiella påverkan.

Denna avhandling syftar till att ge ett ramverk för riskbedömning i AR-utrustade socio-tekniska system genom att identifiera och klassificera mänskliga fel inklusive AR-utökade mänskliga fel och genom att identifiera och klassificera misstag som leder till mänskliga fel, inklusive AR-orsakade misstag. Vårt arbete tillhandahåller också modelleringsfunktioner för socio-tekniska system, för att möjliggöra modellering av AR-relevanta tillförlitlighetshot och för utökade analystekniker för att möta kraven på systemanalys för AR-utrustade socio-tekniska system. För att uppnå detta, föreslår vi en taxonomi för mänskliga funktioner genom att extrahera funktioner från existerande taxonomier över mänskliga fel, organisera och harmonisera dem och sedan utökad taxonomin med AR-utökade funktioner från experiment och studier om förstärkt verklighet. Dessutom föreslår vi en taxonomi av misstag som leder till mänskliga fel genom att extrahera misstag från existerande taxonomier, organisera och harmonisera dem och därefter utöka taxonomin genom att lägga till AR-orsakade misstag från studier och experiment på förstärkt verklighet. I samband av socioteknisk systemmodellering, omvandlas AR-utökade mänskliga funktioner och AR-orsakade fel till förbättrade modelleringselement för både mänskliga och

organisatoriska enheter. SafeConcert, som är en metamodel för modellering av socio-tekniska system, används som grund för utökningen av modelleringselement för socio-enheter. Denna utökade metamodel kan sedan användas för att förbättra de riskanalystekniker som används för socioteknisk systemanalyser. Concerto-FLA, som är en riskanalysteknik för att analysera socio-tekniska system, används som bas för att analysera systembeteende. Vi visar användbarheten av våra modelleringstillägg för akademiska exempel och vi genomför också en fallstudie för att utvärdera analysfunktionerna för de utvecklade utökningarna.

To my family

Acknowledgments

First and foremost, I would like to express my immense gratitude to my main supervisor Barbara Gallina. Thank you for your patience, guidance, encouragement and your perseverance that have inspired me during my research. I also wish to express my gratitude to Karin Laumann and Martin Rasmussen Skogstad from NTNU University, Marko Vidović, Predrag Vidas, Davor Kovačec from Xylon Company, Atanas Gotchev, Robert Bregovic, Minna Luhtanen, Soili Pakarinen from Tampere University for the support and feedback during the ImmerSAFE project.

I also want to take the opportunity to be grateful with the head of our division, Radu Dobrin, as well as Gunnar Widforss, Thomas Nolte, Jan Carlson, Elisabeth Uhlemann, Muhammad Atif Javed and Annika Collander Flytström for facilitating all the MDH routines. My gratitude is also for the people, who are, or have been colleagues at MDH. In particular, I thank Julieth Patricia Castellanos Ardila, Irfan Sljivo, Zulqarnain Haider, for taking their time to answer my questions and sharing their knowledge. Special thanks to Cristina Seceleanu for reviewing my thesis and giving me valuable comments.

Above all, I thank God for helping me in my whole life, then I give special thanks to my parents for always believing in me, offering their most caring support and enthusiasm as well as my family and family-in-law for their inspiration and endless love during these years. Finally, and most important, I would like to express my gratitude and love to my husband Hamed and my son, who will come to this world soon. Their company, unconditional support and love have strengthened me through this challenging experience.

The work in this Licentiate thesis has been supported by EU H2020 MSC-ITN grant agreement No 764951, via the project ImmerSAFE [1].

Soheila Sheikh Bahaei, May, 2020, Västerås, Sweden

List of Publications

Papers Included in the Licentiate Thesis[1]

Paper A: *Augmented Reality-extended Humans: Towards a Taxonomy of Failures - Focus on Visual Technologies*, Soheila Sheikh Bahaei and Barbara Gallina. In Proceedings of the 29th European Safety and Reliability Conference (ESREL-2019), Research Publishing, Singapore, September 2019.

Paper B: *Effect of Augmented Reality on Faults Leading to Human Failures in Socio-technical Systems*, Soheila Sheikh Bahaei, Barbara Gallina, Karin Laumann and Martin Rasmussen Skogstad. In Proceedings of the 4th International Conference on System Reliability and Safety, IEEE, November 2019, indicated as ICSRS-2019a

Paper C: *Extending SafeConcert for Modelling Augmented Reality-equipped Socio-technical Systems*, Soheila Sheikh Bahaei and Barbara Gallina. In Proceedings of the 4th International Conference on System Reliability and Safety, IEEE, November 2019, indicated as ICSRS-2019b.

Paper D: *A Case Study for Risk Assessment in AR-equipped Socio-technical Systems*, Soheila Sheikh Bahaei, Barbara Gallina and Marko Vidović. Technical Report, ISRN MDH-MRTC-332/2020-1-SE, Mälardalen Real-Time Research Center, Mälardalen University, May 2020.

[1] The included papers have been reformatted to comply with the thesis layout

Additional Peer-reviewed Publications Related to the Thesis[2]

Paper 1: *Towards Assessing Risk of Reality Augmented Safety-critical Socio-technical Systems* Soheila Sheikh Bahaei and Barbara Gallina. Published as Proceedings Annex in the 6th International Symposium on Model-Based Safety and Assessment (IMBSA 2019) website, Thessaloniki, Greece, October 2019.

Paper 2: *A Framework for Risk Assessment in Augmented Reality-equipped Socio-technical Systems* Soheila Sheikh Bahaei. Accepted at the Doctoral Forum hosted by the 50th IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2020), Valencia, Spain, June 2020.

[2]These papers are not included in this thesis

Contents

I

Thesis

Chapter 1

Introduction

Augmented reality enhances human performance by expanding human capabilities and upgrading human to an AR-extended human, which is also called augmented human in some literature [2]. For instance, via the usage of visual augmented reality, human vision capabilities may be extended. An example in the automotive domain is the extended situational awareness enabled by adding safety alerts about blind spots of a car on the windshield, which helps a driver to decide comprehensively based on the increased situational awareness [3]. Another example is extending human wayfinding through augmented reality mobile systems by illustrating navigation guidance of landmarks and routes [4].

While the aim of using augmented reality is improving human performance, new types of dependability threats (faults, errors, failures) might be introduced to the system because of these new technologies. In the context of the EU ImmerSAFE project [1], immersive vision-oriented augmented reality, used within safety critical systems, is in focus. Safety critical systems equipped with such augmented reality can be considered as example of socio-technical systems since not only the risk of technical entities has to be assessed in order to ensure safety, but also the risk of non-technical entities such as humans and organizations and effect of augmented reality on them has to be assessed.

If we consider a socio-technical system as a component-based system, then the behavior of the socio-technical system would be the result of the concertation of the various components composing the system: humans, organizations, hardware and software. Based on Avizienis et al. [5]

terminology, any deviation in human functioning from correct functioning is called human failure. However this definition is used for human error in some literature such as [6], we base our work on Avizienis et al. terminology and we call it human failure. In fact, human failure is failure in the last subcomponent of the human composite component. Based on Avizienis et al. terminology, human error is the reason for human failure, which is defined as erroneous human internal state. Fault is the reason for human error, which would be internal or external. Internal faults are originated in internal subcomponents of human component itself and external faults are emanated from other components of the system. For example, social presence can be considered as a subcomponent in human composite component and lack of social presence is an internal fault that would cause failure in human functioning. An experiment on augmented reality [7], shows that using augmented reality would cause diminished social presence, thus using augmented reality introduces this new fault which may cause human failure. Our focus in this study is on human failures and faults leading to human failures. Human error is not considered, because it is related to erroneous human internal state and it can be detected if it leads to human failure. Problems in non-human entities such as technical, environmental and organizational entities in a socio-technical system may cause human failure, thus we consider these problems as external fault category.

Based on ISO 31000: 2018 [8] standard, risk means “effect of uncertainty on objectives” and effect is “deviation from the expected”. Risk is “usually expressed in terms of risk sources, potential events, their consequences and their likelihood”. In most situations it is not possible to provide risk likelihood, because there is not enough experience about those situations for likelihood calculation. New technologies such as augmented reality are in their development process to be used in various industries and there is not statistical information about their utilization to be used for their risk likelihood calculation. Risk sources, potential events and their consequences can be used in risk modeling and analysis methods. We consider three steps for risk assessment, that are identification of dependability threats causing risk, modelling dependability based on identified dependability threats by modelling entities’ behavior that would cause risk and analyzing system behavior to assess risk, shown in Figure 1.1. To do the risk assessment in AR-equipped socio-technical systems, we need to have extension in each of the steps if required, because effect of augmented reality is not considered in current risk assessment techniques. The first step to do the risk assessment, is to identify risk sources and potential events. Our contribution in this step, is

studying effect of augmented reality on human failures and faults leading to human failures, which are considered as risk sources. Fault category includes non-human entities' faults in addition to internal human faults leading to human failures. Second step to do the risk assessment is to model dependability, which means to model entities' behavior and the relationship between them and their consequences. In this step, we consider modelling techniques used for socio-technical systems and how these techniques can be extended to be used for AR applications. Finally, the last step is analyzing the system behavior, which means studying system behavior based on components behavior and their interactions. In this step, we consider an analysis technique used for socio-technical systems and how this technique can be affected by our extensions on modeling. After identifying system behavior, risk control and risk treatment should be done, which means changing the magnitude or likelihood of consequences to increase safety. This step is beyond the scope of this thesis and we consider as our future work.

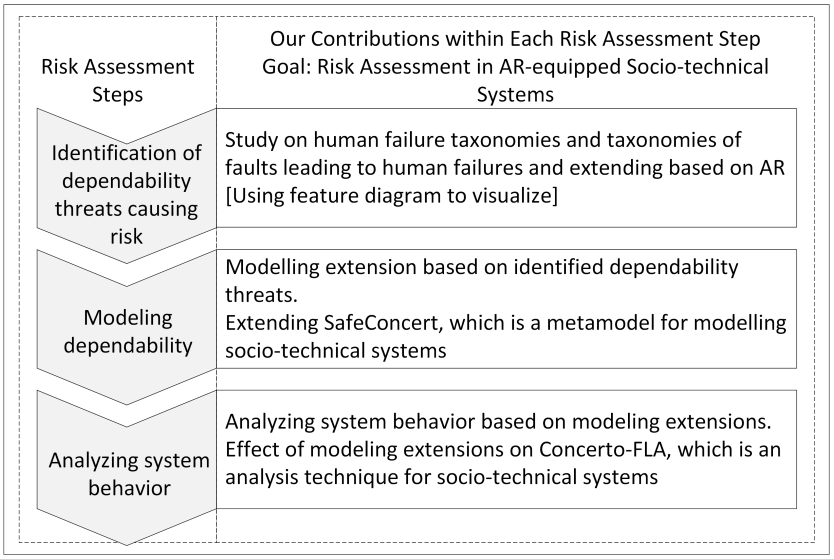

Figure 1.1: Risk assessment steps and our contribution within each step

This thesis aims at providing a framework for risk assessment in augmented reality-equipped socio-technical systems. In particular, we divide uncertainties to two major groups: human functions and other influencing

factors including non-human factors and internal human factors, which can effect on human functions. Deviations in these uncertainties manifest themselves in human failures and faults leading to human failures. In the first step of the risk assessment, we propose a human function taxonomy based on state-of-the-art human failure taxonomies (Norman [9], Reason [10], Rasmussen [11], HFACS (Human Factor Analysis and Classification System) [12], SERA (Systematic Error and Risk Analysis) [13] and Driving [14] human failure taxonomies) and we extend this taxonomy by AR-extended human functions extracted from AR experiments and studies. We also propose a taxonomy of faults leading to human failures based on state-of-the-art taxonomies of faults (Rasmussen [11], HFACS [12], SERA [13], Driving [14] and SPAR-H (Standardized Plant Analysis Risk Human Reliability Analysis) [15]) and we extend it based on experiments and studies on augmented reality. We use feature diagrams to visually illustrate our proposed taxonomies, because it is useful for showing commonalities and variabilities between different taxonomies. To extend modeling step, we use the proposed taxonomies for extending modeling elements in SafeConcert [16], which is a metamodel for modeling component-based socio-technical systems. By extending modelling elements identified dependability threats causing risk can be modelled, which means that their relationship with other entities and their interactions within the whole system can be presented. Finally, for the last step we use the extended metamodel in Concerto-FLA (Failure Logic Analysis) [17], which is an analysis technique for socio-technical systems. Using the analysis technique, system behavior can be identified based on component behavior and their interactions in AR-equipped socio-technical systems.

1.1 Thesis Outline

We organize this thesis in two parts. In the first part, we summarize the research as follows: In Chapter 2, we recall essential background information used throughout this thesis. In Chapter 3, we describe our research methodology and the thesis research goals. In Chapter 4, we describe the specific research contributions of this thesis. In Chapter 5, we discuss related work. Finally, in Chapter 6 we present conclusions and future work.

The second part is a collection of the papers included in this thesis. We now present a brief overview of the included papers.

Paper A: ***Augmented reality-extended humans: towards a taxonomy of failures focus on visual technologies***, Soheila Sheikh Bahaei and Barbara Gallina. In Proceedings of the 29th European Safety and Reliability Conference (ESREL-2019), Research Publishing, Singapore, September 2019.

Abstract: Augmented reality, e.g. immersive visual technologies, augment the human's capabilities. If not properly designed, such augmentation may contribute to the decrease of the human's awareness (e.g., due to distraction) and reaction time efficiency, leading to catastrophic consequences, when included within safety-critical socio-technical systems. Current state-of-the-art taxonomies and vocabularies on human failures do not consider the augmented reality-extended humans. In this paper, first, we review, harmonize and systematically organize the existing human failure taxonomies and vocabularies. More specifically, we consider the existing taxonomies as a product line and propose a feature diagram (visual specification of product lines), which includes the human's functions and the potential failures of those functions, and where commonalities and variabilities represent the evolution over time. Then, to deal with immersive visual technologies, we make the diagram evolve by including additional features. Our feature diagram-given taxonomies of taxonomies may serve as the foundation for failure logic-based analysis of image-centric socio-technical systems.

My contribution: I was the main author of the paper under the supervision of the co-author. My specific contributions included the categorization of the state-of-the-art taxonomies and extracting the extended features based on studies and experiments on AR. Both authors contributed equally in discussions and developing the paper contribution. The co-author contributed with reviews and comments for providing the paper and suggestions/ideas on how to accomplish the task and suggesting to use the feature diagram to have a better visualization.

Paper B: ***Effect of Augmented Reality on Faults Leading to Human Failures in Socio-technical Systems***, Soheila Sheikh Bahaei, Barbara Gallina, Karin Laumann and Martin Rasmussen Skogstad. In Proceedings of the 4th International Conference on System Reliability and Safety, IEEE, November 2019, indicated as ICSRS-2019a.

Abstract: With the ultimate purpose of assessing risk within augmented reality-equipped socio-technical systems, in our previous work, we systematically organized and extended state-of-the-art taxonomies of human failures to include the failures related to the extended capabilities enabled by AR technologies. The result of our organization and extension was presented in form of a feature diagram. Current state-of-the-art taxonomies of faults leading to human failures do not consider augmented reality effects and the new types of faults leading to human failures. Thus, in this paper, we develop our previous work further and review state-of-the-art taxonomies of faults leading to human failures in order to: 1) organize them systematically, and 2) include the new faults, which might be due to AR. Coherently with what done previously, we use a feature diagram to represent the commonalities and variabilities of the different taxonomies and we introduce new features to represent the new AR-caused faults. Finally, an AR-equipped socio-technical system is presented and used to discuss about the usefulness of our taxonomy.

My contribution: I was the main author of the paper. My contribution included the categorization of the state-of-the-art taxonomies and extracting the extended features based on studies and experiments on AR. The co-authors contributed with reviews and comments for improving the paper.

Paper C: ***Extending SafeConcert for Modelling Augmented Reality-equipped Socio-technical Systems***, Soheila Sheikh Bahaei and Barbara Gallina. In Proceedings of the 4th International Conference on System Reliability and Safety, IEEE, November 2019, indicated as ICSRS-2019b.

Abstract: With the emergence of new technologies such as augmented reality in socio-technical systems, traditional risk assessment methods may fail to have a comprehensive system modeling, because these technologies extend human's capabilities, which might introduce new types of human failures caused by failing these extended capabilities and new types of faults leading to human failures. Current state-of-the-art modeling techniques do not contemplate these capabilities and augmented reality-caused faults leading to human failures. In our previous work, we proposed an extension for modeling safety-critical socio-technical systems, to model augmented reality-extended humans by using a taxonomy that contains AR-specific human's failure behavior. In this paper, we continue our extension by investigating faults leading to human failures including faults because of augmented reality. Our

extension builds on top of a metamodel for modeling socio-technical component-based systems, named SafeConcert. We illustrate our extension on two fictitious but credible systems taken from air traffic control and rail industry. In order to model augmented reality-equipped socio-technical systems, we need to consider human and organization as parts of the system and augmented reality as a technology used in the system.

My contribution: I was the main author of the paper. My contribution included extension of the modeling elements based on our proposed taxonomies. The co-author contributed with reviews and comments for improving the paper and suggestion for the basis metamodel for the extension.

Paper D: ***A Case Study for Risk Assessment in AR-equipped Socio-technical Systems***, Soheila Sheikh Bahaei, Barbara Gallina and Marko Vidović. Technical Report, ISRN MDH-MRTC-332/2020-1-SE, Mälardalen Real-Time Research Center, Mälardalen University, May 2020.

Abstract: Augmented Reality (AR) technologies are used as human-machine interface within various types of safety-critical systems. In order to avoid unreasonable risk, it is required to anticipate new types of dependability threats (faults, errors, failures), which could be introduced within the systems by these technologies. In our previous work, we have designed an extension for CHESS framework to capture AR-related dependability threats (focusing on faults and failures) and we have extended its metamodel, which provides qualitative modeling and analysis capabilities that can be used for AR-equipped socio-technical systems. In this paper, we conduct a case study from automotive domain to present modeling and analysis capabilities of our proposed extensions. We conduct qualitative modeling and analysis based on Concerto-FLA analysis technique, which is an analysis technique for socio-technical systems to find out if the proposed extensions would be helpful in capturing new system failures caused by AR-related dependability threats.

My contribution: I was the main author of the paper. My contribution included using of the extensions for the proposed case study. The co-authors contributed with reviews and comments for improving the paper and they provided suggestions for selecting the case and validating the work and information for modeling the case study.

Chapter 2

Background

This section introduces the background required by the current research, helping in the understanding of its content. Section 2.1 provides fundamental definitions for risk assessment and dependability. Our main goal is the provision of a framework for risk assessment in AR-equipped socio-technical systems, thus Section 2.2 provides an overview of augmented reality and Section 2.3 provides an overview of risk assessment in socio-technical systems including essential background related to current human failure taxonomies, modeling dependability and analyzing system behavior. Since our identified dependability threats are presented as feature diagrams, Section 2.4 introduces feature diagrams.

2.1 Fundamental Definitions for Risk Assessment and Dependability

In this subsection, we recall essential definitions related to risk assessment and dependability that will be used during the research.

Based on the definition provided by Aven, risks are "consequences and uncertainties" [18] and risk analysis is a "tool for dealing with uncertainty". Lowrance defines risk as a measure of probability and severity of adverse effects [19]. Based on ICH (International Conference on Harmonization) guidelines [20], risk assessment consists of risk identification, risk analysis and risk evaluation. Risk analysis deals with assigning likelihood and severity to identified risks and evaluation deals with comparing the identified and

analyzed risks against risk criteria to determine whether residual risk is tolerable. Our study is based on ISO 31000: 2018 [8] standard, which is a generic approach and in not related to specific industry. Based on this standard, risk means "effect of uncertainty on objectives" and effect is "deviation from the expected". Risk is "usually expressed in terms of risk sources, potential events, their consequences and their likelihood". Currently, we are focusing on qualitative risk modelling and analysis techniques and we still do not incorporate likelihood in assessment process.

Since we provide examples and case studies from automotive domain, we recall fundamental concepts and standards in this context.

ISO 26262 [21] is a functional safety standard addressing electrical and electronic systems within road vehicles. This standard provides the requirements and set of activities that should be performed during the lifecycle phases such as development, production, operation, service and decommissioning. According to ISO 26262 [21] standard, risk is "combination of the probability of occurrence of harm and the severity of that harm". Risk assessment, which is also called hazard analysis, is a "method to identify and categorize hazardous events of items and to specify safety goals and ASILs (Automotive Safety Integrity Level) related to the prevention or mitigation of the associated hazards in order to avoid unreasonable risk".

Safety element out of context (SEooC) introduced by ISO 26262, part 10, refers to an element that is not defined in the context of a special vehicle, but it can be used to make an item, which implements functions at vehicle level. SEooC is based on ISO 26262 safety process and information regarding system context such as interactions and dependencies on the elements in the environment should be assumed [22].

SEooC system development contains 4 main steps:

1. (a) Definition of the SEooC scope: assumptions related to the scope, functionalities and external interfaces of the SEooC should be defined in this step.
   (b) Definition of the assumptions on safety requirements for the SEooC: assumptions related to item definition, safety goals of the item and functional safety requirements related to SEooC functionality required for defining technical safety requirements of the SEooC should be defined.
2. Development of SEooC: based on the assumed functional safety requirements, technical safety requirements are derived and then SEooC is developed based on ISO 26262 standard.

3. Providing work products: work products, which are documents that show the fulfilled functional safety requirements, assumptions and requirements on the context of SEooC, are provided.

4. Integration of the SEooC into the item: safety goals and functional safety requirements defined in item development should match with assumed functional safety requirements for the SEooC. In case of a SEooC assumption mismatch, change management activity based on ISO 26262 standard should be conducted.

SAE standard [23] describes the taxonomy and definitions related to driving automation systems for on-road motor vehicles performing part or all of the dynamic driving task (DDT) on a sustained basis. Based on this taxonomy, there are six levels of driving automation. SAE level 0 refers to no driving automation and SAE level 5 refers to full driving automation. Assessing human factor in driver-vehicle interface is not only important on lower SAE levels, but also on higher levels because of the importance of safe transition between automated and non-automated vehicle operation [24]. In order to improve safety, various scenarios of driver/vehicle interaction should be considered.

Based on dependability terminology provided by Avizienis et al. [5]:

- *System* is "an entity that interacts with other entities, i.e. other systems, including hardware, software, humans, and the physical world with its natural phenomena".

- *System function* is "what the system is intended to do".

- *Correct service* "is delivered when the service implements the system function".

- *Service failure* or failure is "an event representing a transition (a deviation) from correct service to incorrect service" (shown in Figure 2.1).

- *Human failure* is deviation from correct human function to incorrect human function.

- *Error* "is the part of the total state of the system that may lead to its subsequent service failure" (shown in Figure 2.1).

- *Fault* is "the adjudged or hypothesized cause of an error" (shown in Figure 2.1). It would be internal, if it is emanated from system itself or external, if it is emanated from other systems.

- *Failure modes* is a form that a failure may manifest itself in that form. In literature [25], service's failure modes have been categorized based on:

  1. *Provisioning*
     - *Omission*: No output is provided.
     - *Commission*: Output is provided when not expected.
  2. *Timing*
     - *Early*: Output is provided too early.
     - *Late*: Output is provided too late.
  3. *Value*
     - *Coarse*: The output is not within the expected range of values and user can detect this deviation.
     - *Subtle*: The output is not within the expected range of values and user can not detect this deviation.

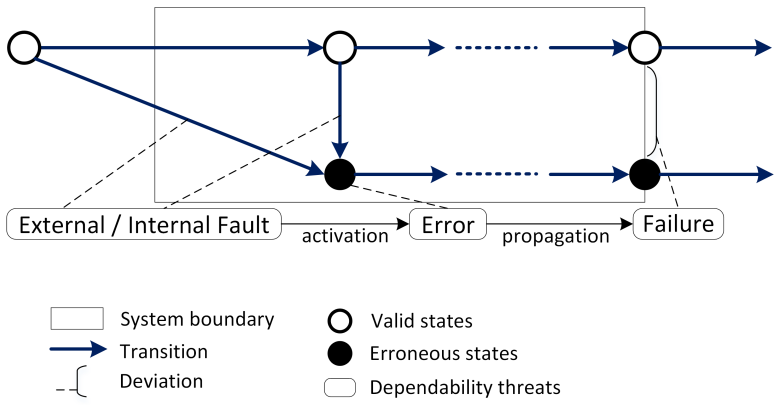

Figure 2.1: Causality chain among threats (adapted from [26])

2.2 Augmented Reality

Augmented reality is any kind of extra information superimposed to reality and provided to user [27]. It would be visual, haptic, auditory, etc. For example, visual augmented reality refers to using graphics and digital content to juxtapose with what an individual is seeing in real-time [28]. However our research is not limited to visual augmented reality, we use visual augmented reality as an example throughout the research, because it is more apprehensible. AR displays can be categorized to three types including head-worn, hand-held and spatial. Head-worn displays are attached to the head, hand-held displays are displays that can be used by hand like mobile phones and spatial displays are placed in the environment like head-up displays (HUDs) [29]. HUD is "any transparent display that presents data without requiring users to look away from their usual viewpoint" [30]. For example, Figure 2.2 shows an example of using augmented reality information illustrating navigation information with the aim of increasing driving efficiency and driver reaction time.

Figure 2.2: Using AR on head up display to show navigation information [31]

Using augmented reality can improve user awareness and reaction time efficiency, meanwhile it can increase cognitive-processing or distract the driver [32], if it covers important parts of the real world view of the driver.

In [33], augmented reality is used in a driver simulator study with 88 participants and results show that visual warnings increase driver performance. Augmented reality can contribute to treatment of several mental and physical disorders [34] and for jobs with demanding situations and

repetitive tasks, which threaten mental and physical health, AR can be used to upkeep mental and physical healthy state [35]. Neurological effects of AR, earned by brain-imaging technology show that brain cognitive activity increases and memory encoding is 70% higher while using AR [36]. AR integrates elements from virtual reality with elements from real world [37] leading to improvement in training by providing interactive ways for engaging learners and motivating them to have a better experience through the augmented environment [38].

Augmented reality may introduce new types of dependability threats. For example, if the expected improvement is not gained through AR because of distracting the user. AR effects on interpersonal communications and decreases social presence [7], which would lead to risk.

2.3 Risk Assessment in Socio-technical Systems

Socio-technical systems are systems including socio entities such as human, organization and technical entities such as software and hardware. Risk assessment in these systems requires identifying dependability threats related to human functioning and influencing factors on human functioning including organizational factors, in addition to dependability threats related to software and hardware. In order to identify dependability threats related to human functions, we had a review on current human failure taxonomies which are briefly introduced in Subsection 2.3.1. Subsection 2.3.2 recalls essential background related to modelling dependability and SafeConcert, which is a metamodel used for modelling socio-technical systems. We can extend this metamodel to model identified human functions and influencing factors. Finally, Subsection 2.3.3 introduces analyzing system behavior and Concerto-FLA, which is an analysis technique for analyzing socio-technical systems and can be used for analyzing AR-equipped socio-technical systems. The reason that we choose this technique for extension is that similar to our work this technique is based on a human failure taxonomy and it models human based on a special taxonomy. We also aim at modelling human based on human failure taxonomies, but we use more taxonomies to cover various human failures and we also extend the taxonomies based on AR experiments and studies.

2.3.1 Human Failure Taxonomies

Based on several studies on causal factors of maritime and aviation accidents such as an analysis on maritime accidents in United States and Canada between 1996 and 2006 [39], human and organizational factors are the most important causal factors for accidents [40]. Thus, to identify dependability threats, we need to find human failure taxonomies to classify them and finally to extend them based on AR effects.

There are several human failure taxonomies providing a taxonomy of human failures in various context. In what follows, we recall the most popular ones. Reason [10] and Norman [9] are two examples of human failure taxonomies. Rasmussen [11] provides a taxonomy including human failures in industrial installations based on analyzing mental processes. HFACS (Human Factor Analysis and Classification System) [12] provides human failure taxonomy based on avionic context, which is based on analysis on 300 aviation accidents. SERA (Systematic Error and Risk Analysis) [13] is another human failure taxonomy, which is developed as a tool for Canadian forces version of HFACS, but it can be used independent of HFACS. Driving failure taxonomy [14] is another taxonomy based on literature on human failures in road transport using dominant psychological mechanisms involved. In this study, we review all mentioned taxonomies and harmonize them to have a general and organized taxonomy which is not specific to special domain and can be used as the basis for modeling and analyzing system behavior in various industries.

2.3.2 Modelling Dependability

There are different modeling languages in the literature to model dependability by proposing UML (Unified Modeling Language) extensions [41]. EAST-ADL2 [42] extends UML and SysML (System Modeling Language) [43] and provides modeling language for automotive domain. ASILs (Automotive Safety Integrity Levels) are used for defining integrity level. DAM (Dependability Analysis Modeling) [44] also provides dependability modeling on UML profile, which is coupled with MARTE (Modeling and Analysis of Real Time and Embedded systems) [45]. We base our work on SafeConcert metamodel [16] because of the support this metamodel provides for modeling socio-technical systems and also because it is integrated within the AMASS platform [46], the first open-source platform for supporting engineering and certification of safety-critical systems [47].

SafeConcert [16] is a metamodel for modeling socio and technical entities in socio-technical systems. This metamodel is part of CHESS ML (CHESS Modeling Language) [48], which is a UML-based modeling language used in CHESS framework [49]. In SafeConcert metamodel, software, hardware and socio entities can be modelled as components in component-based systems representing socio-technical systems. SERA taxonomy [13] is used for modeling human and organization, which are the socio entities of the system.

Human components are represented as composite components and subcomponents are based on twelve categories of human failures in SERA taxonomy. SafeConcert modeling elements are divided into two types based on human functionalities (Figure 2.3). Modeling elements based on functionalities responsible for acting (HumanActuatorUnit) including: *selection*, *response*, *knowledge decision*, *time management*, *communication*, *intent* and *feedback* are shown with prefix "HA" and modeling elements based on functionalities responsible for sensing (HumanSensorUnit) including: *perception*, *attention*, *sensory* and *knowledge perception* are shown with prefix "HS".

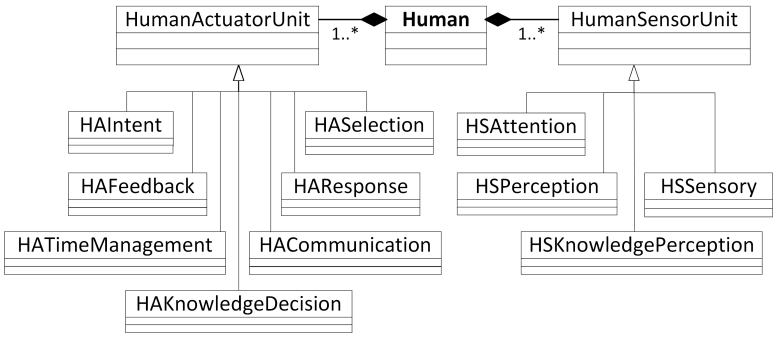

Figure 2.3: SafeConcert modeling elements to model human components [16]

Organization components are represented as composite components and subcomponents are based on six categories of SERA taxonomy. SafeConcert organization modeling elements are shown in Figure 2.4. These subcomponents which are called units in this metamodel are named with prefix "OU" to represent organizational unit.

Based on SafeConcert metamodel, failures are propagated from/to entities in a socio-technical system through ports, and failure modes are associated to ports. Failure modes are assigned to ports by defining failure mode groups

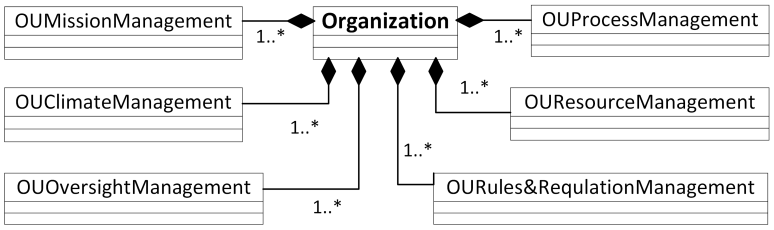

Figure 2.4: SafeConcert modeling elements to model organization components [50]

based on domain [16].

SafeConcert is implemented in CHESS toolset [51] developed within CHESS [52] and Concerto [53] projects and has been included within recently released, open-source AMASS platform for certification [54]. This toolset offers modelling and analysis capabilities targeting high integrity systems as well as socio-technical systems. Users/architects can model the functional view of the architecture of the system as well as the non-functional view (e.g., dependability) of the architecture of the system. Users can define component-based architectural models composed of hardware, software, human and organization. This toolset supports SafeConcert metamodel and can be extended based on the extensions provided for SafeConcert.

2.3.3 Analyzing System Behavior

In socio-technical systems the output is the result of human and technology interaction embedded within social structures such as organizational goals and environmental aspects. Standard techniques in risk analysis such as Fault Tree Analysis (FTA) [55], Failure Modes and Effects Analysis (FMEA) [56], formal methods and probabilistic safety analysis are not sufficient [40]. The problem with traditional methods such as FTA and FMEA is that they should be done manually, which requires a huge amount of time and work for recent complicated systems. Model-driven risk analysis techniques such as Fault Propagation and Transformation Calculus (FPTC) [57], Failure Propagation and Transformation Analysis (FPTA) [58], Hierarchically Performed Hazard and Operability Studies (HiP-HOPS) [59], CHESS-FLA [60] (Failure Logic Analysis within the CHESS project [52]) and Concerto-FLA [17] (Failure Logic Analysis within the Concerto project [53]) are developed based on

traditional methods to automatically provide FTA and FMEA results based on system architecture and modeling of components failure behavior.

Concerto-FLA [17] is a model-based analysis technique that provides the possibility for analyzing failure behavior of humans and organizations in addition to technical entities by using SERA [13] classification of socio-failures. This approach is provided as a plugin within the CHESS toolset and allows users to define component-based architectural models composed of hardware, software, human and organization and for each component, FPTC (Failure Propagation Transformation Calculus) [57] rules (logical expressions that relate output failures to input failures) are used to model a component's failure behavior.

FPTC syntax for modeling failure behavior at component and connector level is as follows:

behavior = expression+

expression = LHS '->' RHS

LHS = portname'.' bL | portname '.' bL (',' portname '.' bL) +

RHS = portname'.' bR | portname '.' bR (',' portname '.' bR) +

failure = 'early' | 'late' | 'commission' | 'omission' | 'valueSubtle' | 'valueCoarse'

bL = 'wildcard' | bR

bR = 'noFailure' | failure

Failure used in this syntax is the form that a failure may manifest itself, which is called failure mode based on dependability terminology provided by Avizienis et al. [5] (explained in Subsection 2.1).

Wildcard in an input port shows that the output behavior is the same regardless of the failure mode on this input port. noFailure in an input port shows normal behavior.

Components' behavior can be classified as source (if component generates a failure), sink (if component is able to detect and correct input failure), propagational (if component propagates failures received in its input to its output) and transformational (if component transforms the type of failure received in its input to another type in its output) [61].

Based on this syntax, "IP1.noFailure $\rightarrow$ OP1.omission" shows a source behavior and should be read as follows: if the component receives noFailure (normal behavior) on its input port IP1, it generates omission on its output port OP1.

Concerto-FLA analysis technique, which uses FPTC syntax includes five main steps.

1. Modeling architectural elements including software, hardware, human, organization, connectors, interfaces and etc.

2. Using FPTC syntactical rules to model failure behavior at component and connector level. Concerto-FLA has adopted FPTC syntax for modeling failure behavior at component and connector level.

3. Modeling failure modes at system level by injection of inputs.

4. Performing qualitative analysis through automatic calculation of the failure propagations. This step is similar to FPTC technique that system architecture is considered as a token-passing network and set of possible failures that would be propagated along a connection is called tokenset (default value for each tokenset is noFailure, which means normal behavior). In order to obtain system behavior, maximal tokenset is calculated for each connection through a fixed-point calculation.

5. Interpreting the results at system level. Based on the interpretation, decision for changing system design would be taken.

We show these steps by providing an example to clarify how the technique works. HUD system explained in Subsection 2.2 is used as an example.

1. In the first step of Concerto-FLA technique, model of architectural elements should be provided. Based on system description, HUD and human are considered as two composite components of the system. Architectural model of the system using SafeConcert modeling elements is shown in Figure 2.5. HUD is composed of three main elements: combiner, projector unit and computer. Combiner is any transparent display for illustrating AR information. AR information is projected on the combiner by projector unit and is produced by computer [30]. Computer receives raw data from sensors. Because of the presence of AR technology, we call the composite component AR-HUD. To model human composite component, three human modeling elements are selected from Figure 2.3, including HSPerception, HAKnowledgeDecision and HAResponse, which are modeled as three subcomponents of human composite component. Output of the HAResponse component is output of the system, which is shown by human function.

2. In the second step, failure behaviors of each component should be provided using FPTC rules, which are based on studying each

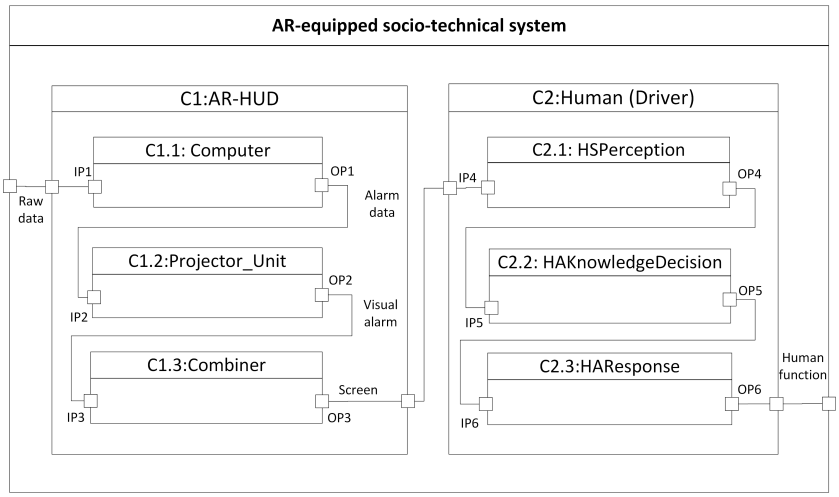

Figure 2.5: Concerto-FLA modelling for AR-HUD example

component in isolation. Incoming and outgoing failures can be classified by related domain failure categorization. For example, timing, value, commission and omission failures are considered in this approach. "IP1.noFailure → OP1.noFailure" behavior shows that if there is normal behavior in input of computer subcomponent, then there is normal behavior in its output. Some sample rules for subcomponents are shown in Figure 2.6.

3. In the third step, we assume that the raw data is provided late by sensor and late will be considered as the input failure for computer subcomponent (IP1 in Figure 2.6).

4. In the fourth step, we calculate the failure propagations, which is shown in Figure 2.7. Based on the analysis algorithm provided in Concerto-FLA technique, each subcomponent is considered as a point and default tokenset is assigned to all connections between subcomponents. Tokenset for each connection is defined with a noFailure token. Then, maximal tokenset is provided based on FPTC expressions and by comparing input failure mode with left hand side of the FPTC expressions. Right hand side of the matched expressions will be added to tokenset of the outgoing connection. For example, possible

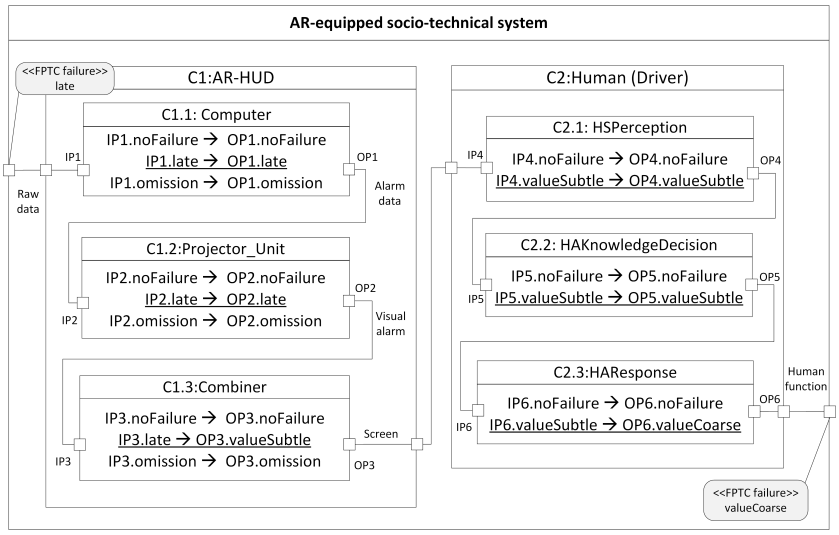

Figure 2.6: Concerto-FLA modelling and analysis results on AR-HUD example

failure modes for IP1 are noFailure and late (noFailure is the default failure mode for all connections and late is shown in the picture as the possible input failure mode). Based on the FPTC expressions in computer component, noFailure and late match with left hand side of the first two expressions, thus their right hand side will be added to IP2 tokenset. The failure propagation is calculated for all connections and maximal tokenset is calculated (shown in Figure 2.7). The failure propagation leads to valueCoarse failure in system output. This step is done automatically in CHESS toolset.

5. Finally, in the last step results can be interpreted. ValueCoarse on OP6 is because of valueSubtle on IP6 that is because of valueSubtle on IP5 and we continue this back propagation to find the origin of the failure that is late on IP1 in this case (shown in Figure 2.8). By using this method, it is possible to find the effect of Components' failure behavior on critical systems' failure behavior considering the origin of the failure. Then, mitigation methods can be used to mitigate the failures and the analysis can be used iteratively to reach the required level of safety.

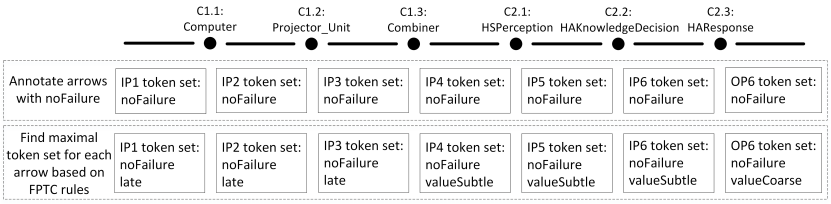

Figure 2.7: Concerto-FLA analysis on AR-HUD example

valueCoarse on OP6 -> valueSubtle on IP6 -> valueSubtle on IP5 -> valueSubtle on IP4 -> late on IP3 -> late on IP2 -> late on IP1

Figure 2.8: Back propagation of the results on AR-HUD example

2.4 Feature Diagrams

In this subsection, we introduce feature diagrams, because we use feature diagrams to illustrate our proposed taxonomies. A feature is a distinguishing attribute of a family of systems, which can be recognized by end-users [62]. Family of products are known as product lines [63]. Feature diagrams are multi-level trees in which nodes are features and edges are used for decomposition of features to more specific features. These diagrams can be used to illustrate common and distinctive features of a product line. Figure 2.9 shows a simple example of a feature diagram. Features can have different types, for example mandatory, optional and alternative [62]. Mandatory features shown by solid dot are essential in the system, which means that all the products in a product line have these features, whereas optional features (a node with a circle) are optional, hence some products may not have those features. Alternative features (XOR) are those features that would not unite in each product of the product line. In this example, a family of visual AR devices are described with display being a mandatory feature, hence all visual AR devices have a display, but remote control and internet connection are optional, as there are some visual AR devices without remote control and internet connections. Display feature would be transparent or nontransparent, hence these are alternative features.

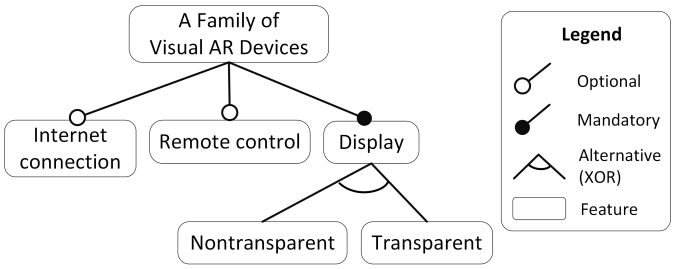

Figure 2.9: Feature diagram of a family of visual AR devices [64]

Chapter 3

Research Summary

In this section, a summary of the research is presented. First, we describe the research methodology applied in Section 3.1, then we present the problem to be solved in Section 3.2 and research goals and subgoals in Section 3.3.

3.1 Research Methodology

Conducting research in a particular area requires comprehending related research methods and being able to apply them. A framework for research methods within computing area is shown in Figure 3.1 [65]. There are four main steps including problem identification, data collection, data processing and evaluating the result. The research starts with *identifying the problem* and defining what we want to achieve and what is happening. This step can be conducted through study of state-of-the-art. The next step is *data collection*, where it is required to define how and where to collect data. This step can be conducted through literature review. Once the data is collected, it should be processed through the step *processing data*. Processing data can be conducted through classifying data and creating taxonomy methodologies. Finally, the last step is *evaluating the result*, where goal achievement can be analyzed and limitations can be identified. Evaluating the results can be conducted through conducting a case study.

An overview of the adapted research method used in this thesis is shown in Figure 3.2. First, we identify the research problem and define the main goal. Then, we divide the research problem to sub-problems and identify the

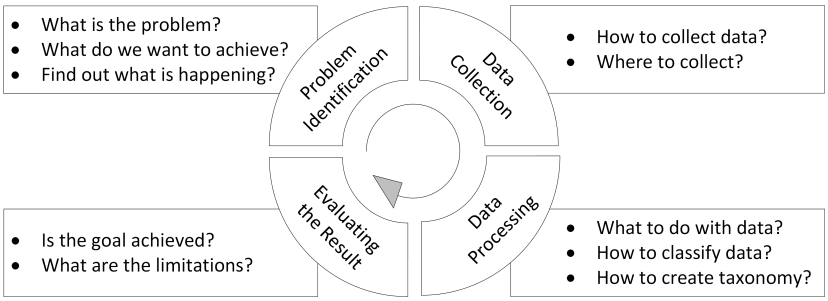

Figure 3.1: A framework for research methods within computing area

subgoals. After that, we propose a solution for the gap identified on the study. Next, we implement the solution and evaluate on an academic example. After this step, the results can be published as a paper. Integration and communication with industry can also be considered as steps after academic evaluation of the proposed solution, to enable evaluating the solution on real world problem. Finally, if the result is accepted for the real world problem, it will provide the possibility of publishing a paper, otherwise problem should be identified to repeat the iterative task for the new research problem.

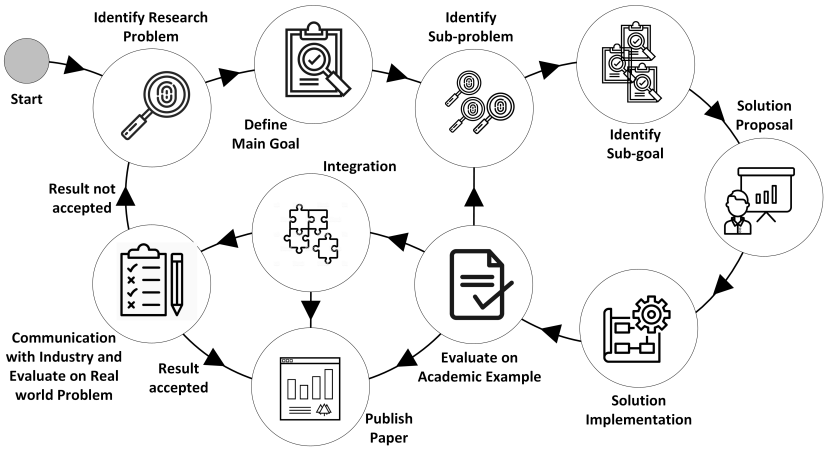

Figure 3.2: Overview of our research methodology

In order to identity research problem and sub-problems we used state of the art review. In order to define main goal and subgoals we used literature review. To propose solutions we used qualitative data analysis and to implement them we used classification and tabulation, to make it possible to have an overview of the collected data. Finally, for evaluation and integration, we used case study methodology.

3.2 Problem Statement

New technologies, such as augmented reality are used with the aim of increasing human performance and extending human capabilities, meanwhile failing of these extended capabilities introduces new types of failures. Thus, these technologies may cause new types of human failures and new types of faults leading to human failures. In socio-technical systems, system contains socio entities including human and organization and technical entities including software and hardware. The first step to do the risk assessment, is to identify dependability threats. Next steps are to provide modeling and analysis means for dependability information. There are various human failure taxonomies and taxonomies of influencing factors on human failures that can be used as the basis in analysis tools. However, there are no data on effect of augmented reality and the new types of human failures and faults leading to human failures that would be introduced to the system because of augmented reality.

This thesis aims at providing a framework for risk assessment in augmented reality-equipped socio-technical systems. More specifically, the thesis identifies the effect of augmented reality on human failures and faults leading to human failures to get involve modelling elements and analysis tools based on these effects. To reach this goal, first state-of-the-art human failure taxonomies and taxonomies of faults leading to human failures are studied, categorized and extended based on experiments and studies on augmented reality, then these taxonomies are used for extending modeling elements used in SafeConcert, which is a metamodel for modeling socio-technical systems. Last but not least, we also consider effect of these extensions on analysis techniques for socio-technical systems such as Concerto-FLA, to be able to analyze AR-equipped socio-technical systems.

3.3 Research Goals

As presented in Section 3.2, this thesis aims at providing a framework for risk assessment in augmented reality-equipped socio-technical systems. To reach this target, we define the main research goal as follows:

Overall Research Goal:
Assessing risk in augmented reality-equipped socio-technical systems.

In order to address the overall research goal, we define concrete subgoals that address the specific challenges. Subgoals are based on three steps defined for risk assessment in Figure 1.1. The subgoals are described as follows:

Subgoal 1: ***Identifying and classifying the common and variable human-related dependability threats in relation to the technological and organizational changes.*** AR-extended human failures and faults leading to human failures including AR-caused faults are considered as human-related dependability threats in AR-equipped socio-technical systems. For each of the human failures and fault categories a taxonomy is provided. Providing an AR-extended human function taxonomy and a taxonomy of faults leading to human failures for socio-technical systems, requires a study on current state-of-the art taxonomies. There are two challenges in this research subgoal. The first challenge is that there are different taxonomies with various categorizations on human failures and faults leading to human failures, thus a systematic organization on different taxonomies is essential to reach a harmonized taxonomy. The second challenge is that effect of augmented reality is not considered in taxonomies, thus a review on experiments and studies on augmented reality is required to evolve the harmonized human function taxonomy based on the extended human capabilities and to evolve harmonized taxonomy of faults leading to human failures based on augmented reality-caused faults. Once dependability threats are identified based on an AR-extended human function taxonomy and a taxonomy of faults leading to human failures, the next step is to propose representation means for the identified dependability threats.

Subgoal 2: ***Developing representation means for capturing the behavior of the involved entities and the behavioral result of their interactions***

within AR-equipped socio-technical systems. We extend a tool supported metamodel for modelling socio-technical systems, to enable modelling of identified dependability threats and representing their relationship through defining modelling elements. Once representation means are developed, the next step is to analyze system behavior to assess risk in AR-equipped socio-technical systems.

Subgoal 3: ***Analyzing the behavior of AR-equipped socio-technical systems such that risk can be assessed.*** In order to meet this goal, it is required to know how the analysis would be done based on current analysis techniques for socio-technical systems using the extensions and proposed representation means. Then, based on the results required extensions can be proposed for the current analysis techniques.

Chapter 4

Thesis Contributions

In this chapter, we present a brief description of the technical contributions provided by this thesis. In particular, in Section 4.1, we describe the first contribution, which is a taxonomy of AR-extended human functions based on state-of-the-art human failure taxonomies and studies and experiments of augmented reality. We named this taxonomy AREXTax. In Section 4.2, we describe the second contribution, which is a taxonomy of faults leading to human failures including AR-caused faults based on state-of-the-art fault taxonomies and studies and experiments of augmented reality. We named this taxonomy AREFTax. These two taxonomies are used for representing the behavior of the involved entities in AR-equipped socio-technical systems. In Section 4.3, we describe the third contribution, which is our proposed representation means by extending modeling elements in an existing metamodel for modeling AR-equipped socio-technical systems based on the identified dependability threats. Finally, in Section 4.4, we describe the fourth contribution, which is analysis of AR-equipped socio-technical systems behavior using AR-extensions.

4.1 Augmented Reality-extended Human Function Taxonomy

AREXTax on AR-extended humans shown in Figure 4.1, is a human function taxonomy based on state-of-the-art human failure taxonomies including Norman [9], Reason [10], Rasmussen [11], HFACS [12], SERA [13] and

Driving [14]. This taxonomy is illustrated via a feature diagram (recalled in Subsection 2.4), to visually show the commonalities and variabilities of different categorizations. The taxonomy is extended for augmented reality-extended humans by adding AR-extended human functions as extended features.

In this taxonomy, human functions are extracted from human failure taxonomies. For example, paying attention function is extracted from attention failure. AR-extended functions are shown by dotted lines border rectangles. For example, surround detecting is an extended function which is added because of using augmented reality [66]. For example, AR information on the windshield of a car showing blind spots, helps a driver to detect surrounding environment, which is an AR-extended function [30].

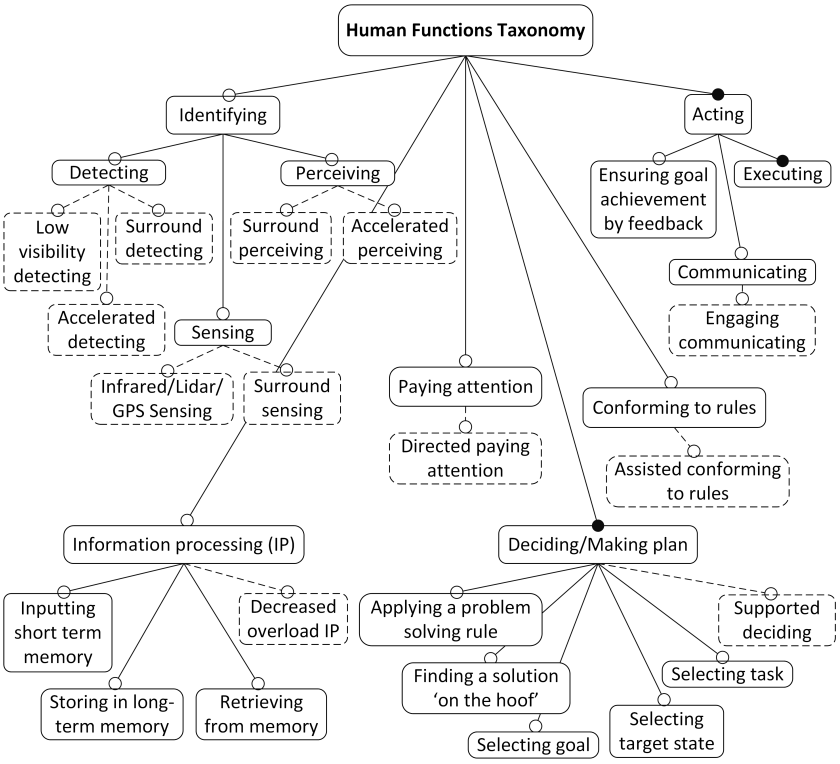

Figure 4.1: AR-extended human function taxonomy [64]

This taxonomy can be used as a list of AR-extended human functions in an AR-equipped socio-technical system. List of functions are required for safety risk assessment, because in risk assessment techniques components are defined based on functions and then possible failures of functions should be considered while analyzing the components. Common and variable functions extracted from various state-of-the-art human failure taxonomies are shown using the feature diagram. For example, deciding is extracted from decision failure and this failure is common in six above mentioned taxonomies. This taxonomy is proposed in paper *"Augmented reality-extended humans: towards a taxonomy of failures – focus on visual technologies"*.

4.2 Taxonomy of Faults Leading to Human Failures

AREFTax on faults leading to human failures including AR-caused faults is shown in Figure 4.2, which is a taxonomy based on state-of-the-art fault taxonomies including Rasmussen [11], HFACS [12], SERA [13], Driving [14] and SPAR-H [15]. This taxonomy is illustrated as a feature diagram, to visually show the commonalities and variabilities of different categorizations. The taxonomy is extended for augmented reality-caused faults leading to human failures by adding AR-caused faults as extended features based on experiments and studies on augmented reality. For example, social faults (problems in communicating with others) are personnel faults, categorized based on state-of-art taxonomies, which might lead to human failures. Using augmented reality may decrease social presence and a new type of fault (social presence fault) may lead to human failures [7]. This new type of fault as AR-caused faults are shown by dotted border rectangular in the taxonomy.

This taxonomy can be used as a list of identified faults leading to human failures in an AR-equipped socio-technical system, which shows the common and variable features between various state-of-the-art taxonomies of faults leading to human failures. For example, environment fault is common in all five taxonomies. This taxonomy is proposed in the paper *"Effect of augmented reality on faults leading to human failures in socio-technical systems"*.

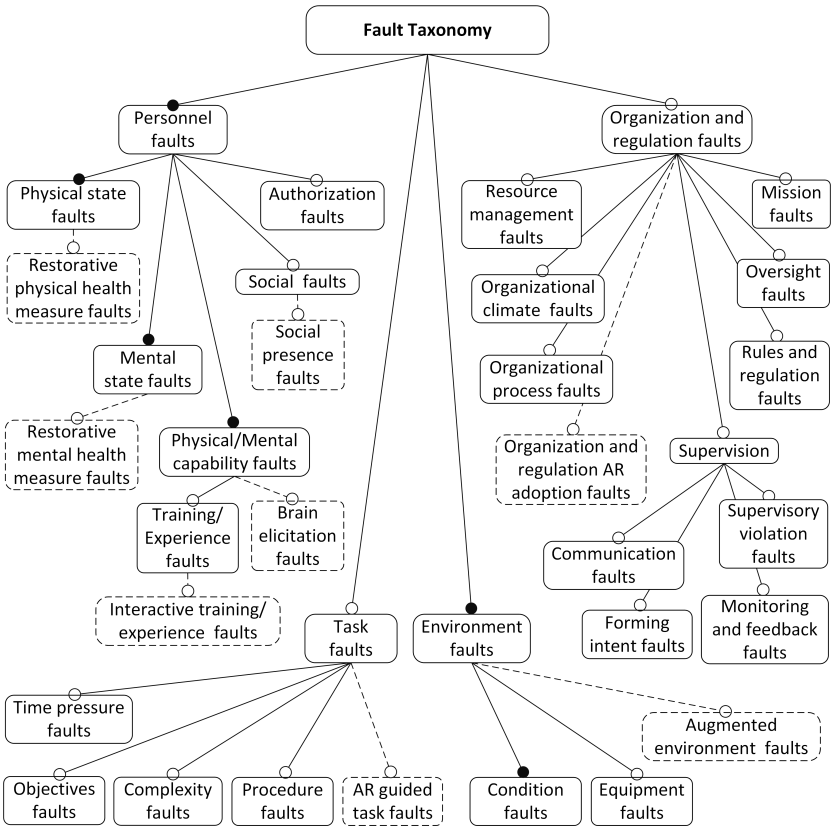

Figure 4.2: Taxonomy of faults leading to human failures including AR-caused faults [31]

4.3 Representation Means for Modeling AR-extended Humans and AR-caused Faults

SafeConcert is a metamodel for modeling socio-technical systems (recalled in Subsection 2.3.2). Based on the human function taxonomy proposed in Subsection 4.1 and the taxonomy proposed in Subsection 4.2, we extend human and organization modeling elements to empower analysis tools based on this metamodel, to model augmented reality-extended humans and

augmented reality-caused faults leading to human failures. There are six categories of human functions (shown in Figure 4.1), which can be grouped in three types of human functionalities. Functions for earning situational awareness including identifying and paying attention, functions for information processing and deciding and functions for acting and conforming to rules. These categories are shown by HumanSAUnit, HumanProcessUnit and HumanActuatorUnit (shown in Figure 4.3). Modeling elements characterizing AR-extended human functions are shown with dotted line border rectangles and commonalities with SafeConcert are shown by grey color, to clarify the extensions. We also extend human modeling elements based on the personnel faults in fault taxonomy (shown in Figure 4.2), and consider a human fault unit including these internal faults. The outcome is shown in Figure 4.3.

In addition, we extend organization modeling elements by considering organization, task and environment faults leading to human failures. These faults are shown in Figure 4.2. The extended modeling elements based on these faults are shown in Figure 4.4. Modeling elements characterizing AR-caused faults are shown in dotted line border rectangles and commonalities with SafeConcert are shown by grey color, to clarify the extensions.

The extensions for AR-extended human modeling elements and AR-caused faults modeling elements based on AREXTax and AREFTax are presented in the paper *"Extending SafeConcert for Modelling Augmented Reality-equipped Socio-technical Systems"*.

We use the extended SafeConcert for modelling the example shown in Figure 2.2. We can consider three composite components including human component, organization component (road transport organization) and AR-HUD component. We consider organization component also to take into account effect of organizational factors (such as environmental factors). Organizational factors influencing human functioning are selected from extended SafeConcert organization modeling elements shown in Figure 4.4. Our selected elements are:

- Organization and regulation AR adoption: it refers to upgrading rules and regulations of road transport organization based on AR technology [67].
- Condition: it refers to road condition.
- AR guided task: it refers to the task, which AR is used for guiding driver

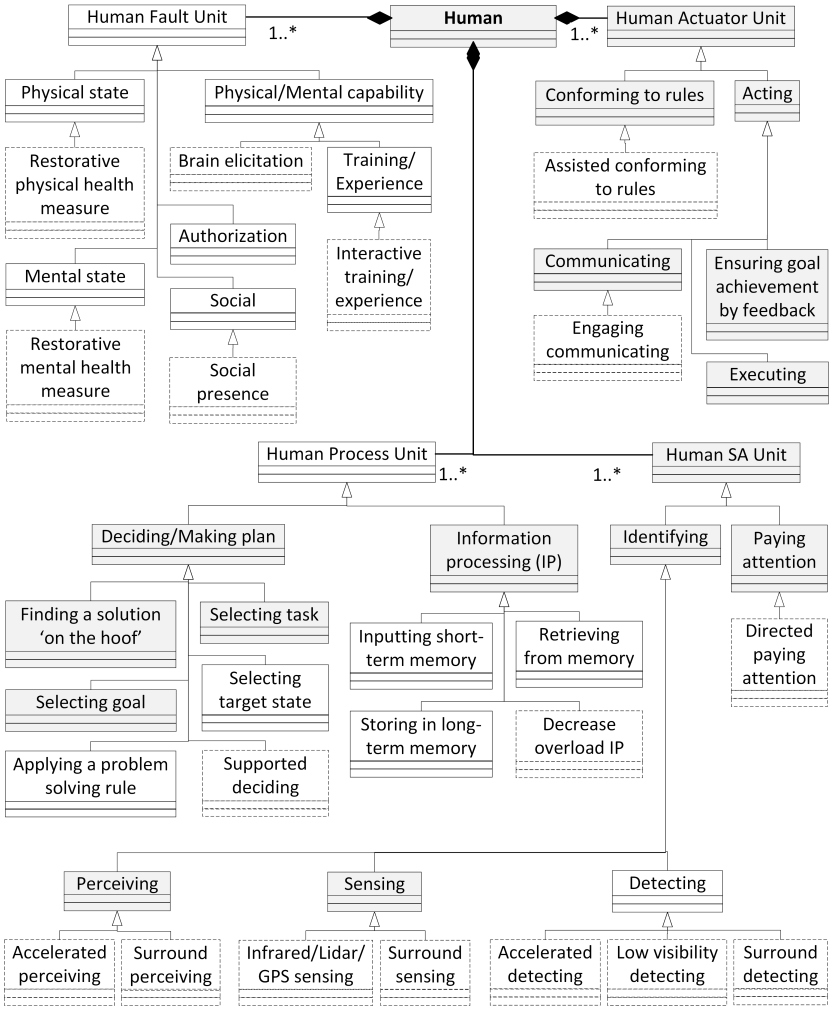

Figure 4.3: Extended SafeConcert human modeling elements [50]

to do that [68]. For example, if AR is used to guide driver to park the car more safely, parking safely is the AR-guided task.

Organization component receives input from system, which represents

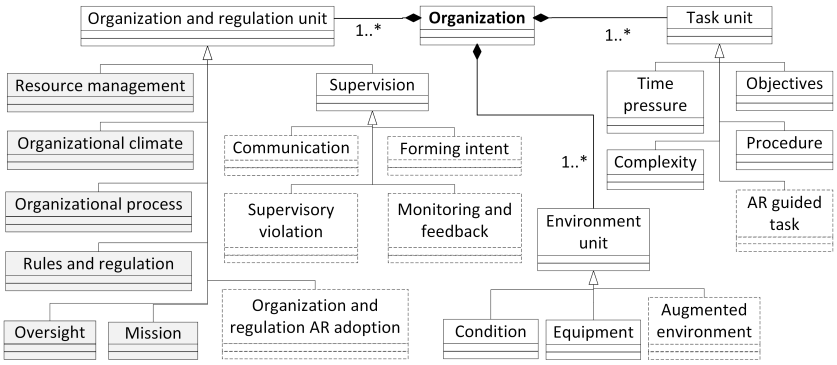

Figure 4.4: Extended SafeConcert organization modelling elements [50]

influences from regulation authorities on the organization (REG).

We consider four subcomponents of human composite component selected from extended SafeConcert human modeling elements shown in Figure 4.3. These four subcomponents are:

- Surround detecting: it refers to an AR-extended function, because driver can detect surround environment through AR technology.
- Deciding: it refers to human decision making function.
- Executing: it refers to human executing function.
- Social presence: it refers to an AR-caused factor, because AR may decrease social presence and lead to human failure.

Surround detecting effects on deciding and deciding effects on executing. Social presence input is connected to system input with the name human communication input (HCI) and effects on human executing. Human output, which is output of the system is human function shown by HF.

An AR-HUD component contains three primary subcomponents [30]:

- Projector unit: it refers to the subcomponent that produces an image on a combiner.
- Combiner: it refers to the subcomponent that is a flat piece of glass and can be the windshield of the car.

- Computer: it refers to the subcomponent that generates the information that should be displayed by projector unit.

Another system input, which is input of the computer subcomponent is raw data (RD) provided by sensors.

To illustrate the case study, we explain about three scenarios depicted in Figure 4.5. AR-extended function and AR-caused faults are shown by gray color, to show the effect of AR and the contribution of the proposed modelling elements.

In the first scenario (S1), content provided by AR-HUD is wrong and it leads to the driver's failure. For example, there is failure in combiner of AR-HUD, which is a technical component. This failure is an external fault for human component and causes system failure.

In the second scenario (S2), content provided by AR-HUD is correct, but there is failure in organization and regulation AR adoption, which is an external fault for human component. For example, when the organization does not provide facilities for using AR in organization, then the organization does not provide required condition and does not define the guiding task using AR. This failure is also an external fault for human component and causes system failure.

In the third scenario (S3), there is failure in social presence subcomponent of the driver component, which is an internal fault leading to failure in executing subcomponent and leading to system failure. For example, driver would miss the common ground with other people, this failure would lead to wrong action. Social presence is an internal fault for human component and would lead to system failure.

As it is shown in this example, the proposed extended modeling elements can be used for enhancing modelling of internal and external faults leading to human failures, used in risk analysis tools.

4.4 Analysis of AR-equipped Socio-technical System Behavior Using AR-extensions

In order to analyze AR-equipped socio-technical systems, we need to decorate the architectural elements including AR-extended elements with dependability information concerning their failure behavior. We use Concerto-FLA as the basis of our work, since this analysis technique is proposed for socio-technical systems. It is implemented in CHESS framework [49] and contains language

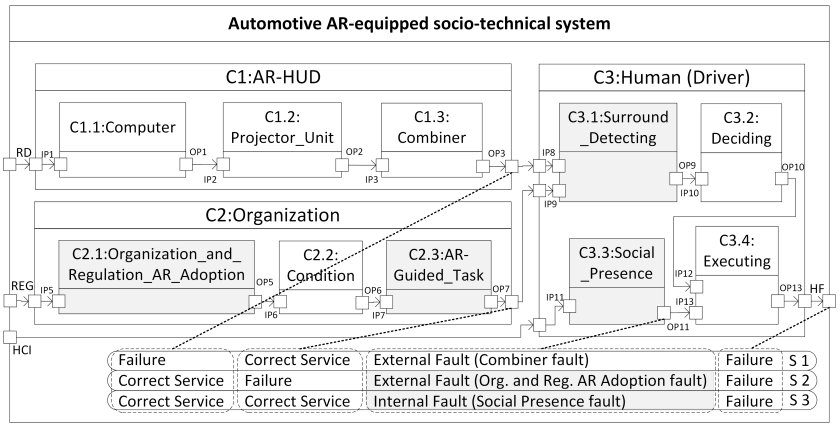

Figure 4.5: Using extended SafeConcert for modelling an AR-equipped socio-technical system

constructs for human and organization components based on the SafeConcert metamodel. It is required to extend these language constructs to be able to use dependability information related to AR-extended elements while doing the analysis.

In paper *"A Case Study for Risk Assessment in AR-equipped Socio-technical Systems"*, we conducted an industrial case study from automotive domain to present analysis capabilities provided by using the AR-extensions. Case study is based on a surround view system as a safety element out of context (SEooC) defined in ISO 26262, which is selected in cooperation with our industrial partners.

In this subsection, we use the HUD example and present the analysis capabilities provided by using AR-extensions on this example. We use five steps of the Concerto-FLA analysis technique explained in Subsection 2.3.3 to present analysis results and discuss about effect of AR-extensions.

1. First step is provided in Figure 4.5. We explained how the system is modeled in Subsection 4.3.

2. Second step is shown by providing FPTC rules, which is used for linking possible failures of inputs of each component to failures of outputs. (For the sake of brevity, we show two rules of each subcomponent in Figure 4.6)

3. Third step is considering possible failures in inputs of the system. In this example, we inject noFailure to three inputs of the system, because we aim at analyzing system for scenarios that failure is emanated from our modeled system.

4. Fourth step is calculating the failure propagations. We consider a scenario and show the analysis results in Figure 4.6.

5. Last step is back propagation of results (Shown in Figure 4.7). Interpretation of the back-propagated results can be used to make decision about design change or defining safety barrier, if it is required.

In the scenario shown in Figure 4.6, we assume that the road transport organization has not updated rules and regulations based on AR technology. Therefore, this component will produce an omission failure. We show the failure propagation with underlined FPTC rules, which are the rules that are activated, shown in Figure 4.6. In this scenario, AR-HUD sub-components behave as propagational and propagate noFailure from input to output. Organization and regulation AR adoption behaves as source and while its input is noFailure, it produces omission failure in its output. The activated rule is underlined on this component. Omission failure propagates through condition and AR guided task and in surround detecting it transforms to valueSubtle. The reason for this transformation is that omission failure in IP9 means that AR guided task is not defined by organization. This means that surround detecting would be done incorrectly, because its input is not provided and this leads to valueSubtle failure in its output. ValueSubtle propagates to deciding and transforms to valueCoarse in executing. The reason for this transformation is that if there is value failure in executing function it can be detected by user, which means valueSubtle transforms to valueCoarse.

Based on back propagation of the results, shown in Figure 4.7, we can explain how the rules have been triggered. ValueCoarse on OP13 is because of valueSubtle on IP12. ValueSubtle on IP12 is because of valueSubtle on IP10 and we continue this back propagation to reach a component originating the failure, which is component with input IP5 that is organization and regulation AR adoption. In this case, a solution would be an instruction for organization and regulation to update their rules and regulations based on AR technology. Then, the failure behavior will be updated and failure propagation analysis can be repeated for another iteration.

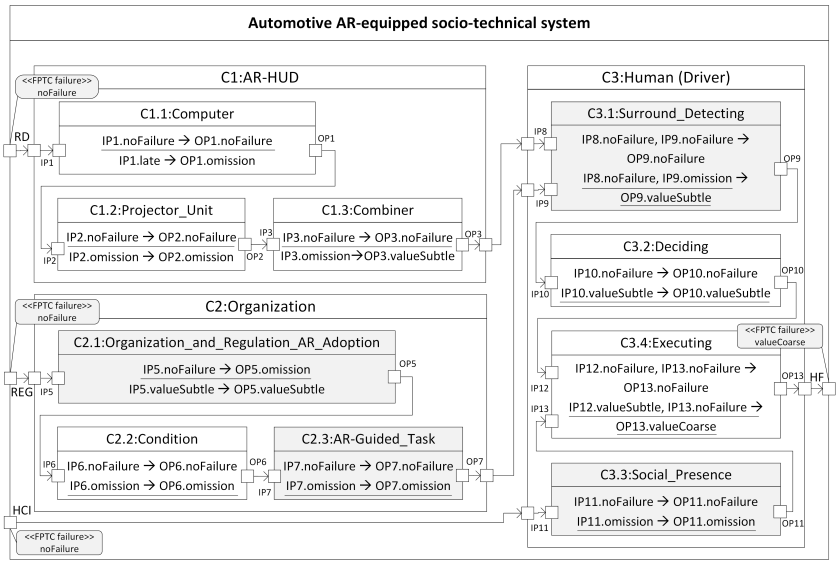

Figure 4.6: Analyzing AR-equipped socio-technical system using AR-extensions

valueCoarse on OP13 -> valueSubtle on IP12 -> valueSubtle on IP10 -> omission on IP9 -> omission on IP7 -> omission on IP6 -> noFailure on IP5

Figure 4.7: Back propagation of the results on AR-HUD example using AR-extensions

As it is shown in this case study, by using modeling elements related to AR-extended human functions as well as modeling elements related to AR-caused faults leading to human failures and by analyzing their failure propagation, architects and safety managers have at disposal means to reveal effect of AR-related dependability threats on system behavior. For example, in this scenario, it is not possible to detect risks emanated from failure in updating rules and regulations based on AR technology, without using the proposed representation means. Using these representation means or modeling elements provides the possibility to analyze their failure propagation and provides the possibility to analyze effect of these failures on system behavior. Then, based on analysis results decision about design change or fault mitigation mechanisms would be taken.

Chapter 5

Related Work

In this section, related work is discussed. In Section 5.1, works that address modeling of socio-technical systems are presented. In Section 5.2, works that address risk analysis in socio-technical systems are presented.

5.1 Modeling Socio-technical Systems

In [69], the authors propose a technique for modeling global software development project as a complex socio-technical system. In this method, functional components are identified and links between the components are defined. Feedback controller is used between two components to control if there is any deviation between the interpretation of the component providing the output and the component receiving the output as its input. Feedback controller implementation can not be done through mechanical device and informal communication is required. The modeling technique provided in this paper is specific for software development as a socio-technical system and can not be used for other domains. In comparison, we extend modeling techniques that can be used for socio-technical systems including hardware, software and socio entities used in various domains.

In [70], a safety risk framework with the name Socio-Technical Risk Analysis (SoTeRiA) is proposed, which provides a theoretical basis for integration of technical system risk models with social and structural aspects of models for safety prediction. In particular, this method extends PRA (Probabilistic Risk Analysis) framework [71] to add organizational aspects.

This study is similar to our work in that it considers human, software, hardware and organizational failures and provides the organizational safety causal model. The difference is that we use the list of human failures and influencing factors to provide the modeling elements and the analyzer can choose the related elements for the specified case.

In [72], authors propose a modelling methodology for complex socio-technical systems while new technologies are used by humans. In this method, technology modelling is used to consider its impact on system's behavior and it consists of CWA (Cognitive Work Analysis) [73] and SD (System Dynamics) [74] approaches to capture effect of humans and dynamic interactions in complex systems. The difference of this work from ours is that the focus in this work is on complex socio-technical systems for systems engineering.

In [75], the author proposes SD-BBN, which is a method that combines Bayesian belief networks (BBN) [76] and system dynamics (SD) [74] for socio-technical predictive modeling. In BBN, probabilities of causes and effects are shown by conditional probabilities. Expert opinion is used for defining the probabilities. To consider feedback loops and dynamic interactions of causal factors, this method combines BBN with SD. SD is a simulation-based modeling technique that is useful for modeling organizational behavior, dynamics and feedback. This SD-BBN method is integrated with classical probabilistic risk analysis (PRA) [71] techniques and fault tree and event tree are used to model system risk. This model is used to predict happening of accidents in a period of time and guide managers to schedule their activities, while our model is used during the system development process for eliminating design failures incrementally and iteratively.

In [77], the author proposes an accident model for socio-technical systems based on system theory called STAMP (Systems-Theoretic Accident Model and Processes). In this model, the focus is on continuous controlling and defining safety constraints to keep system with safe behavior while changes and adaptations instead of focusing on preventing component failures. Thus, examination should be done in all levels of socio-technical systems to identify each level's contribution to the loss. To accomplish this goal, identifying the factors leading to accident is required. STAMP uses accident reports to identify the required information. However, for developing system engineering techniques to prevent accidents, these information are not at disposal.

5.2 Risk Analysis in Socio-technical Systems

As it was explained in Chapter 1, modeling can be considered as part of risk analysis and we model AR-equipped socio-technical systems to empower analysis techniques to do risk analysis in these systems. In some of the works risk analysis is done based on questionnaires and ratings provided by people using the system. For example, in [78], risk analysis for context-adaptive augmented reality aerodrome control towers assistance system is done through ratings provided by aerodrome controllers using the system in a simulation environment. Criteria used for risk analysis are transparency, complexity, interference, disruptiveness, distraction potential, failure modes and trust/complacency. The results of the analysis show that this system is supportive for air traffic controllers and provides safety benefits. This study would be useful for demonstrating the effectiveness of using augmented reality in aerodrome control towers assistance systems. Instead, our approach includes modeling of the AR-equipped socio-technical systems to analyze safety and find the design and implementation problems during the system development.

In [79], an initial solution for risk analysis in safety-critical applications using augmented reality is proposed, which is named Safe-AR. This method integrates risk analysis with three phases at which AR interacts with the user. These three phases are perception, comprehension and decision making. Safe-AR integrate failure modes associated with above mentioned three phases of user's mental information processing into the risk analysis. Four failure modes are considered for each phase. In our approach, instead, we consider several state-of-the-art taxonomies to reach a comprehensive risk identification on human and influencing factors and then we extend the taxonomies based on AR experiments and studies to identify AR-extended human failures and AR-caused faults leading to human failure.

Chapter 6

Conclusions and Future Work

In this chapter, we first summarize our work and provide concluding remarks and then we present the future research directions.

6.1 Conclusions

The goal of our research is to provide a framework for risk assessment in augmented reality-equipped socio-technical systems. For achieving our goal, we focused on various kinds of dependability threats that would cause risk and providing means for modeling and analyzing system behavior in order to be able to assess those risks. We defined three subgoals (presented in detail in Section 3.3):

- **Subgoal 1**: Identifying and classifying the common and variable human-related dependability threats in relation to the technological and organizational changes.
- **Subgoal 2**: Developing representation means for capturing the behavior of the involved entities and the behavioral result of their interactions within AR-equipped socio-technical systems.
- **Subgoal 3**: Analyzing the behavior of AR-equipped socio-technical systems such that risk can be assessed.

To reach the specified subgoals, we presented set of research contributions (detailed in Chapter 4):

- **Thesis contribution 1**: AREXTax, a taxonomy of AR-extended human functions.
- **Thesis contribution 2**: AREFTax, a taxonomy of faults leading to human failures including AR-caused faults.
- **Thesis contribution 3**: Representation Means for Modeling AR-extended Humans and AR-caused Faults.
- **Thesis contribution 4**: Analysis of AR-equipped socio-technical system behavior Using AR-extensions.

Figure 6.1 presents the mapping between subgoals (described in Chapter 3, Section 3.3), research contributions (described in Chapter 4) and included papers (described in Chapter 1. Section 1.1).

6.1.1 Research Subgoal 1

The aim of using new technologies such as augmented reality in socio-technical systems is to increase human performance. However, these technologies would introduce new dependability threats to the system that should be considered in risk assessment. To support capturing human-related dependability threats in socio-technical systems, human failure taxonomies can be used. However, it is important to consider effect of technological and organizational changes on human behavior.

To support capturing human-related dependability threats in relation to technological and organizational changes, we proposed an AR-extended human function taxonomy that allows capturing human behaviors extended by using augmented reality. This taxonomy is based on state-of-the-art human failure taxonomies and by categorizing their commonalities and variabilities. In addition, we considered influencing factors on human behaviors that failing in their functioning would act as fault for human behavior. We proposed a fault taxonomy including AR-caused faults leading to human failures, to support capturing of influencing factors' effect on human functioning and also effect of augmented reality on them. This taxonomy is also based on state-of-the-art fault taxonomies and by categorizing their commonalities and variabilities.

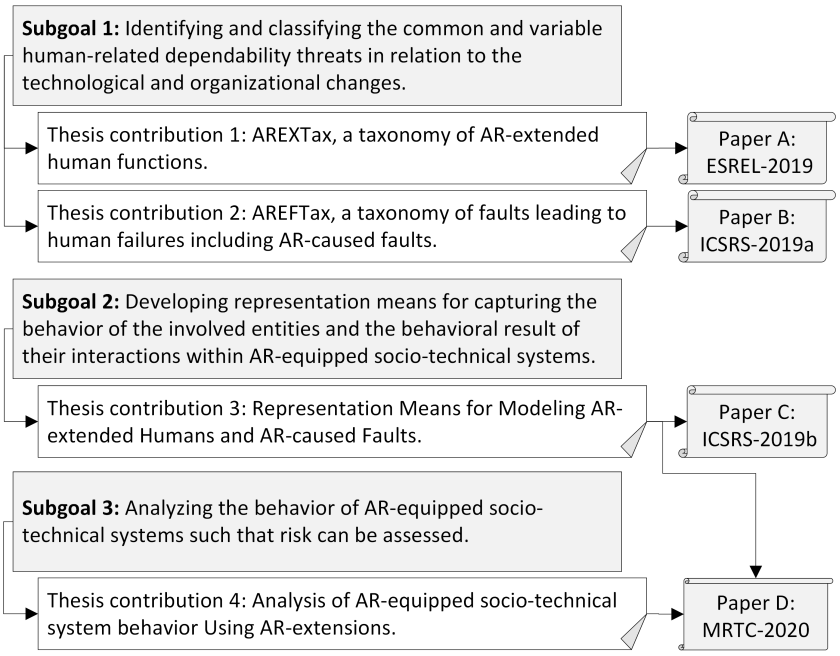

Figure 6.1: Connection between subgoals, contributions and the papers

6.1.2 Research Subgoal 2

There are modeling languages and constructs for representing involved entities' behavior. To support capturing new dependability threats in AR-equipped socio-technical systems, it is required to provide representation means for capturing behaviors and their interaction within new technologies. We designed extensions for a metamodel aiming for socio-technical systems, to provide extended modeling elements that can model effect of augmented reality. In order to reach this goal, human modeling elements are extended to be able to model AR-extended human functions and organization modeling elements are extended to model non-human influencing factors on human behavior.

6.1.3 Research Subgoal 3

In order to assess the risk of identified dependability threats, after modeling involved entities' behavior, it is required to analyze system behavior. There are various methods for analyzing system behavior, based on entities' behavior. We use Concerto-FLA, because it is an analysis technique for socio-technical systems. We use this technique by considering our extended modeling elements in order to evaluate analysis capabilities provided by our proposed extensions. We use an industrial case study to show the analysis results.

6.2 Future Work

The contributions provided in this thesis can be improved in several directions. Here we present the suggested areas for future work.

- Classification of human functions in the proposed AR-extended human function taxonomy is based on the studied taxonomies and by considering their commonalities and variabilities. It is not mutually exclusive when classifying human functions using this taxonomy, because it is possible to have situations that a human failure can be categorized in more than one single category. There is a need to work more on the classification to omit the overlap between the provided categories. However, it requires more research on psychology and human factors.

- Classification of faults leading to human failures in the proposed fault taxonomy is based on the studied taxonomies and by considering their commonalities and variabilities. It is not mutually exclusive when classifying faults using this taxonomy, because it is possible to have situations that a fault can be categorized in more than one single category. Thus, a possible direction of improvement would be to work more on the classification to omit the overlap between the provided categories.

- The reason that we decided to work on SafeConcert metamodel and to extend this metamodel is that this metamodel provides modeling elements for socio-technical systems and also because it is integrated within the AMASS platform, the first de-facto open-source platform for supporting engineering and certification of safety-critical cyber physical

systems. One research direction is to have a systematic literature review on different metamodels and modeling languages to be able to use the extensions in other metamodels and to use their advantages in our extensions.

- The reason that we decided to work on Concerto-FLA analysis technique and to extend this technique is that this technique provides analysis means for socio-technical systems and also it is integrated in the AMASS platform. One research direction is to have a systematic literature review on different socio-technical analysis techniques to be able to use the extensions in other methods and to use their advantages in our extensions.

- The proposed extension for extending CHESS toolset is not implemented in this toolset. We aim at implementing the conceptual extension of SafeConcert within CHESSML, in order to make evolve the analysis plugin of CHESS toolset.

- Implementing the conceptual extensions of CHESSML provides the possibility for implementing the extensions on analysis, in order to have the analysis results automatically. We aim at extending Concerto-FLA based on extended modeling elements to provide the analysis extension required in the third step of the risk assessment process within the CHESS toolset.

- We conducted a case study based evaluation on our extensions. However, because of the low criticality of the selected case, we could not model and analyze high risk scenarios. In future, we aim at using industrial cases with higher criticality and we aim at using the implemented technique to analyze dependability of AR-equipped socio-technical systems.

- In this thesis, we focused on identifying the dependability threats and assessing risk caused by these dependability threats. We did not provide any mitigation technique for the identified risks and this can be considered as future work to decrease risk and provide risk management techniques.

Bibliography

[1] ImmerSAFE.: Immersive Visual Technologies for Safety-critical Applications. https://immersafe-itn.eu (2019)

[2] Raisamo, R., Rakkolainen, I., Majaranta, P., Salminen, K., Rantala, J., Farooq, A.: Human augmentation: Past, present and future. International Journal of Human-Computer Studies (IJHCS) **131** (2019) 131 – 143

[3] Abdi, L., Abdallah, F.B., Meddeb, A.: In-vehicle augmented reality traffic information system: a new type of communication between driver and vehicle. Procedia Computer Science **73** (2015) 242–249

[4] Goldiez, B.F., Ahmad, A.M., Hancock, P.A.: Effects of augmented reality display settings on human wayfinding performance. Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) **37**(5) (2007) 839–845

[5] Avizienis, A., Laprie, J.C., Randell, B., Landwehr, C.: Basic concepts and taxonomy of dependable and secure computing. IEEE Transactions on dependable and secure computing (TDSC) **1**(1) (2004) 11–33

[6] Fastenmeier, W., Gstalter, H.: The human reliability approach in road traffic safety. In: European Conference on Human Centred Design for Intelligent Transport Systems. Retrieved from http://www. conference2010. humanist-vce. eu. (2010)

[7] Miller, M.R., Jun, H., Herrera, F., Villa, J.Y., Welch, G., Bailenson, J.N.: Social interaction in augmented reality. PloS one **14**(5) (2019) e0216290

[8] International Organization for Standardization (ISO). : ISO 31000: Risk management – Guidelines. (2018)

[9] Norman, D.A.: Errors in human performance. Technical report, California Univ San Diego LA JOLLA Center For Human Information Processing (1980)

[10] Reason, J.: The human contribution: unsafe acts, accidents and heroic recoveries. CRC Press (2017)

[11] Rasmussen, J.: Human errors. a taxonomy for describing human malfunction in industrial installations. Journal of occupational accidents **4**(2-4) (1982) 311–333

[12] Shappell, S.A., Wiegmann, D.A.: The human factors analysis and classification system–HFACS. Technical report, Civil Aeromedical Institute (2000)

[13] Hendy, K.C.: A tool for human factors accident investigation, classification and risk management. Technical report, Defence Research And Development Toronto (Canada) (2003)

[14] Stanton, N.A., Salmon, P.M.: Human error taxonomies applied to driving: A generic driver error taxonomy and its implications for intelligent transport systems. Safety Science **47**(2) (2009) 227–237

[15] Gertman, D., Blackman, H., Marble, J., Byers, J., Smith, C., et al.: The SPAR-H human reliability analysis method. US Nuclear Regulatory Commission **230** (2005)

[16] Montecchi, L., Gallina, B.: SafeConcert: A metamodel for a concerted safety modeling of socio-technical systems. In: International Symposium on Model-Based Safety and Assessment (IMBSA), Springer (2017) 129–144

[17] Gallina, B., Sefer, E., Refsdal, A.: Towards safety risk assessment of socio-technical systems via failure logic analysis. In: International Symposium on Software Reliability Engineering Workshops (ISSRE), IEEE (2014) 287–292

[18] Aven, T.: Foundations of risk analysis. John Wiley & Sons (2012)

[19] Lowrance, W.W.: Of Acceptable Risk: Science and the Determination of Safety. (1976)

[20] Guideline, I.H.T.: Quality risk management. Q9, Current step **4** (2005) 408

[21] International Organization for Standardization (ISO). : ISO 26262: Road vehicles — Functional safety. (2018)

[22] Gallina, B., Carlson, J., Hansson, H., et al.: Using safety contracts to guide the integration of reusable safety elements within ISO 26262. In: 21st Pacific Rim International Symposium on Dependable Computing (PRDC), IEEE (2015) 129–138

[23] Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles: https://www.sae.org/standards/content/j3016_201806/ (2018)

[24] Dimitrakopoulos, G., Uden, L., Varlamis, I.: The Future of Intelligent Transport Systems. Elsevier (2020)

[25] Pumfrey, D.J.: The principled design of computer system safety analyses. PhD thesis, University of York (1999)

[26] Gallina, B.: PRISMA: a software product line-oriented process for the requirements engineering of flexible transaction models. PhD thesis, University of Luxembourg (2010)

[27] Goldiez, B.F., Saptoka, N., Aedunuthula, P.: Human performance assessments when using augmented reality for navigation. Technical report, University of Central Florida Orlando Inst for Simulation and Training (2006)

[28] Roitman, L., Shrager, J., Winograd, T.: A comparative analysis of augmented reality technologies and their marketability in the consumer electronics segment. Journal of Biosensors and Bioelectronics (JBSBE) **8**(01) (2017)

[29] Van Krevelen, D., Poelman, R.: A survey of augmented reality technologies , applications and limitations. The International Journal of Virtual Reality (IJVR) **9**(2) (2010) 1–20

[30] Phan, M.T.: Estimation of driver awareness of pedestrian for an augmented reality advanced driving assistance system. PhD thesis, Université de Technologie de Compiègne (2016)

[31] Sheikh Bahaei, S., Gallina, B., Laumann, K., Rasmussen Skogstad, M.: Effect of augmented reality on faults leading to human failures in socio-technical systems. In: International Conference on System Reliability and Safety (ICSRS), IEEE (2019)

[32] Hall, N., Lowe, C., Hirsch, R.: Human factors considerations for the application of augmented reality in an operational railway environment. Procedia Manufacturing **3** (2015) 799–806

[33] Schwarz, F., Fastenmeier, W.: Augmented reality warnings in vehicles: Effects of modality and specificity on effectiveness. Accident Analysis & Prevention **101** (2017) 55–66

[34] Ventura, S., Baños, R.M., Botella, C.: Virtual and augmented reality: New frontiers for clinical psychology. State of the Art Virtual Reality and Augmented Reality Knowhow (N Mohamudally, Eds.) Rijeka: InTech (2018) 99–118

[35] Salamon, N., Grimm, J.M., Horack, J.M., Newton, E.K.: Application of virtual reality for crew mental health in extended-duration space missions. Acta Astronautica **146** (2018) 117–122

[36] Heather, A.: How augmented reality affects the brain. Technical report, Neuro-Insight (2018)

[37] Gutiérrez, M., et al.: Augmented reality environments in learning, communicational and professional contexts in higher education. Digital Education Review **26** (2014) 22–35

[38] Lee, K.: Augmented reality in education and training. TechTrends **56**(2) (2012) 13–21

[39] Johnson, C.W., Holloway, C.M.: A longitudinal analysis of the causal factors in major maritime accidents in the USA and Canada (1996–2006). In: The Safety of Systems. Springer (2007) 85–104

[40] Qureshi, Z.H.: A review of accident modelling approaches for complex socio-technical systems. In: Proceedings of the twelfth Australian workshop on Safety critical systems and software and safety-related programmable systems-Volume 86, Australian Computer Society, Inc. (2007) 47–59

[41] Booch, G., Rumbaugh, J., Jackobson, I.: UML: Unified Modeling Language (1997)

[42] Berntsson, L.O., Blom, H., Chen, D., Cuenot, P., Freund, U., Frey, P., Gérard, S., Johansson, R., Lönn, H., Reiser, M.O., et al.: EAST-ADL2 UML2 Profile specification (2008)

[43] Friedenthal, S., Moore, A., Steiner, R.: OMG systems modeling language (OMG SysML) tutorial. In: INCOSE Intl. Symp. Volume 9. (2006) 65–67

[44] Bernardi, S., Merseguer, J., Petriu, D.C.: A dependability profile within MARTE. Software & Systems Modeling **10**(3) (2011) 313–336

[45] André, C., Cuccuru, A., Dekeyser, J.L., De Simone, R., Dumoulin, C., Forget, J., Gautier, T., Gérard, S., Mallet, F., Radermacher, A., et al.: MARTE: a new OMG profile RFP for the modeling and analysis of real-time embedded systems. In: DAC 2005 Workshop-UML for SoC Design. (2005)

[46] AMASS open platform: https://www.polarsys.org/opencert/news/2018-12-05-download_p2_preview/ (2018)

[47] AMASS – Architecture-driven, Multi-concern and Seamless Assurance and Certification of Cyber-Physical Systems.: http://www.amass-ecsel.eu/ (2018)

[48] CONCERTO D2.7 – Analysis and back-propagation of properties for multicore systems – Final Version: http://www.concerto-project.org/results (2016)

[49] Mazzini, S., Favaro, J.M., Puri, S., Baracchi, L.: CHESS: an Open Source Methodology and Toolset for the Development of Critical Systems. In: EduSymp/OSS4MDE@ MoDELS. (2016) 59–66

[50] Sheikh Bahaei, S., Gallina, B.: Extending safeconcert for modelling augmented reality-equipped socio-technical systems. In: International Conference on System Reliability and Safety (ICSRS), IEEE (2019)

[51] Cicchetti, A., Ciccozzi, F., Mazzini, S., Puri, S., Panunzio, M., Zovi, A., Vardanega, T.: CHESS: a model-driven engineering tool environment for aiding the development of complex industrial systems. In: Proceedings of the 27th IEEE/ACM International Conference on Automated Software Engineering, ACM (2012) 362–365

[52] ARTEMIS-JU-100022 CHESS – Composition with guarantees for high-integrity embedded software components assembly: http://www.chess-project.org/ (2010)

[53] ARTEMIS-JU-333053 CONCERTO – Guaranteed Component Assembly with Round Trip Analysis for Energy Efficient High-integrity Multi-core systems: http://www.concerto-project.org (2016)

[54] Ruiz, A., Gallina, B., de la Vara, J.L., Mazzini, S., Espinoza, H.: Architecture-driven, multi-concern and seamless assurance and certification of cyber-physical systems. In: International Conference on Computer Safety, Reliability, and Security (SafeComp), Springer (2016) 311–321

[55] Vesely, W.E., Goldberg, F.F., Roberts, N.H., Haasl, D.F.: Fault tree handbook. Nuclear Regulatory Commission Washington DC (1981)

[56] Stamatis, D.H.: Failure mode and effect analysis: FMEA from theory to execution. Quality Press (2003)

[57] Wallace, M.: Modular architectural representation and analysis of fault propagation and transformation. Electronic Notes in Theoretical Computer Science **141**(3) (2005) 53–71

[58] Ge, X., Paige, R.F., McDermid, J.A.: Probabilistic failure propagation and transformation analysis. In: International Conference on Computer Safety, Reliability, and Security (SafeComp), Springer (2009) 215–228

[59] Papadopoulos, Y.: Safety-directed system monitoring using safety cases. PhD thesis, Citeseer (2000)

[60] Gallina, B., Javed, M.A., Muram, F.U., Punnekkat, S.: A model-driven dependability analysis method for component-based architectures. In: 38th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), IEEE (2012) 233–240

[61] Šljivo, I., Gallina, B., Carlson, J., Hansson, H., Puri, S.: A method to generate reusable safety case argument-fragments from compositional safety analysis. Journal of Systems and Software **131** (2017) 570–590

[62] Kang, K.C., Cohen, S.G., Hess, J.A., Novak, W.E., Peterson, A.S.: Feature-oriented domain analysis (FODA) feasibility study. Technical

report, Carnegie-Mellon Univ Pittsburgh Pa Software Engineering Inst (1990)

[63] Schobbens, P.Y., Heymans, P., Trigaux, J.C., Bontemps, Y.: Generic semantics of feature diagrams. Computer Networks **51**(2) (2007) 456–479

[64] Sheikh Bahaei, S., Gallina, B.: Augmented reality-extended humans: towards a taxonomy of failures – focus on visual technologies. In: European Safety and Reliability Conference (ESREL), Research Publishing, Singapore (2019)

[65] Holz, H.J., Applin, A., Haberman, B., Joyce, D., Purchase, H., Reed, C.: Research methods in computing: What are they, and how should we teach them? In: Working group reports on ITiCSE on Innovation and technology in computer science education. (2006) 96–114

[66] Fu, W.T., Gasper, J., Kim, S.W.: Effects of an in-car augmented reality system on improving safety of younger and older drivers. In: International Symposium on Mixed and Augmented Reality (ISMAR), IEEE (2013) 59–66

[67] Chandra, S., Kumar, K.N.: Exploring factors influencing organizational adoption of augmented reality in e-commerce: Empirical analysis using technology-organization-environment model. Journal of Electronic Commerce Research (JECR) **19**(3) (2018)

[68] Condino, S., Carbone, M., Piazza, R., Ferrari, M., Ferrari, V.: Perceptual limits of optical see-through visors for augmented reality guidance of manual tasks. Transactions on bio-medical engineering (2019)

[69] Bider, I., Otto, H.: Modeling a global software development project as a complex socio-technical system to facilitate risk management and improve the project structure. In: 10th International Conference on Global Software Engineering (ICGSE), IEEE (2015) 1–12

[70] Mohaghegh, Z., Mosleh, A.: Incorporating organizational factors into probabilistic risk assessment of complex socio-technical systems: Principles and theoretical foundations. Safety science **47**(8) (2009) 1139–1158

[71] Bedford, T., Cooke, R., et al.: Probabilistic risk analysis: foundations and methods. Cambridge University Press (2001)

[72] Oosthuizen, R., Pretorius, L.: Assessing the impact of new technology on complex socio-technical systems. South African Journal of Industrial Engineering (SAJIE) **27**(2) (2016) 15–29

[73] Jenkins, D.P., Stanton, N.A., Walker, G.H.: Cognitive work analysis: coping with complexity. CRC Press (2017)

[74] Bayer, S.: Business dynamics: systems thinking and modeling for a complex world. Interfaces **34**(4) (2004) 324–326

[75] Mohaghegh, Z.: Combining system dynamics and bayesian belief networks for socio-technical risk analysis. In: International Conference on Intelligence and Security Informatics (ISI), IEEE (2010) 196–201

[76] Pearl, J.: Bayesian networks: A model of self-activated memory for evidential reasoning. In: Proceedings of the 7th Conference of the Cognitive Science Society. (1985) 329–334

[77] Leveson, N.: A new accident model for engineering safer systems. Safety science **42**(4) (2004) 237–270

[78] Gürlük, H., Gluchshenko, O., Finke, M., Christoffels, L., Tyburzy, L.: Assessment of risks and benefits of context-adaptive augmented reality for aerodrome control towers. In: Digital Avionics Systems Conference (DASC), IEEE (2018) 1–10

[79] Lutz, R.R.: Safe-AR: Reducing risk while augmenting reality. In: 29th International Symposium on Software Reliability Engineering (ISSRE), IEEE (2018) 70–75

II

Included Papers

Chapter 7

Paper A: Augmented Reality-extended Humans: Towards a Taxonomy of Failures – Focus on Visual Technologies

Soheila Sheikh Bahaei, Barbara Gallina
In Proceedings of the 29th European Safety and Reliability Conference (ESREL-2019), Hannover, Germany, September 2019.

Abstract

Augmented reality, e.g. immersive visual technologies, augment the human's capabilities. If not properly designed, such augmentation may contribute to the decrease of the human's awareness (e.g., due to distraction) and reaction time efficiency, leading to catastrophic consequences, when included within safety-critical socio-technical systems. Current state-of-the-art taxonomies and vocabularies on human failures do not consider the augmented reality-extended humans. In this paper, first, we review, harmonize and systematically organize the existing human failure taxonomies and vocabularies. More specifically, we consider the existing taxonomies as a product line and propose a feature diagram (visual specification of product lines), which includes the human's functions and the potential failures of those functions, and where commonalities and variabilities represent the evolution over time. Then, to deal with immersive visual technologies, we make the diagram evolve by including additional features. Our feature diagram-given taxonomies of taxonomies may serve as the foundation for failure logic-based analysis of image-centric socio-technical systems.

7.1 Introduction

Augmented reality-extended humans refers to humans, who can see, hear, perhaps touch, smell and taste more than the non-extended ones by receiving extra information through augmented reality [1]. For example in transport system, additional information regarding surrounding environment can be displayed on the windshield of the car to extend driver capabilities in driving safely [2].

Providing extra information through visual augmented reality can improve driver's performance, but meanwhile it can enforce additional cognitive-processing load [3] or distract driver, if it is not properly designed. Failures related to using visual augmented reality technology or more specifically immersive visual technologies are not considered by current human failure taxonomies. In this paper, first, we review state-of-the-art human failure vocabularies and taxonomies with the lens of the well-established terminological framework on dependability [4]. Then, we provide a novel organization of the fragmented taxonomic domain knowledge by means of a feature diagram that systematizes their inherent commonality and variability. Finally, we extend the feature diagram by considering failures describing the deviating behavior of augmented reality-extended humans, focusing on visual technologies. The final outcome serves as the foundation for failure logic-based analysis tools for (image-centric) socio-technical systems.

The rest of the paper is organized as follows. In Section 7.2, we provide essential background information. In Section 7.3, we review human failure taxonomies, with the state-of-the-art dependability-focused lens. In Section 7.4, we propose our human failure taxonomy. In Section 7.5, we discuss about our achievements. Finally, in Section 7.6, we draw our conclusions and sketch future work.

7.2 Background

In this section, we provide the background information on which this work is based on.

7.2.1 Feature Model and Feature Diagram

A *feature* is a prominent or distinctive characteristic of a family of systems that can be understood or seen by end-users [5]. For example, transmission and horn in a family of bicycles. *Feature modeling* deals with the illustration of common and distinctive features of a family of products. Families of products are also known as *product lines* [6]. *Feature diagrams* are a broadly used specification language for modelling features. A *feature diagram* consists of a multi-level tree, where nodes are features and edges are used to decompose features into more detailed features. There are different kind of features such as mandatory, optional and alternative [5]. The legend in Figure 7.1 summarizes the subset of the concrete syntax of feature diagrams, used in this paper. The feature diagram, in Figure 7.1, exemplifies the usage of feature diagrams for a family of bicycles, characterized by four features, where transmission feature is mandatory, horn is optional. One gear or multi gears, which specialize transmission, are given in alternative.

7.2.2 Basic Concepts on Dependable Systems

In this subsection, we recall essential dependability-related terms, introduced by Avizienis et al. [4]. *System* is "an entity that interacts with other entities, i.e. other systems, including hardware, software, humans, and the physical world with its natural phenomena". *System function* is "what the system is intended to do" and *correct service* "is delivered when the service implements the system function". *Service failure* or failure is "an event representing a transition (a deviation) from correct service to incorrect service." *Error* "is the

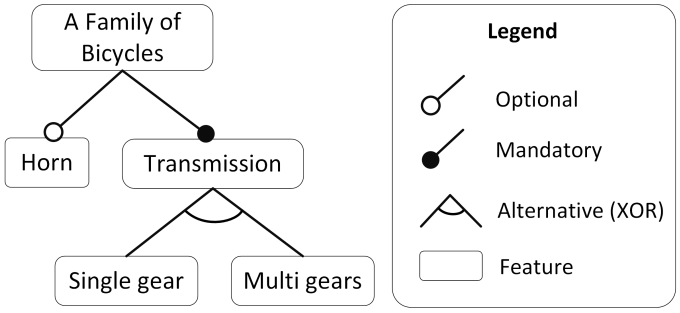

Figure 7.1: Feature diagram of a family of bicycles

part of the total state of the system that may lead to its subsequent service failure". *Fault* is "the adjudged or hypothesized cause of an error". A failure may manifest itself in different forms that are called *failure modes*. In literature [7], service's failure modes have been categorized based on: 1) provisioning (*omission, commission*); 2) timing (*early, late*); 3) value (*course, subtle*).

7.2.3 Visual Augmented Reality Technology

Visual Augmented Reality (AR) technologies [1] superimpose computational and virtual content upon the real world view of the users. We summarize some of the effects of using augmented reality from various research papers:

1. Drivers may detect risks and respond more quickly [8]; detect hazards in low visibility [9].

2. Drivers' perception to side lanes vehicles may be augmented [8] and the drivers' speed in perceiving [10] may be increased.

3. Driver's situation awareness [8] may be augmented. Note that situation awareness is "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" [**?**]. For example, when a pedestrian is in front of the car, the driver first perceives the pedestrian, than estimates the time for crossing (comprehend) and then decides about the action (projection). Therefore, increased situation awareness shows improvement in perceiving and deciding functions.

4. In visual augmented reality technologies, GPS, lidar and infrared sensors provide more information from outside of the vehicle for driver and extend human sensing/ detecting/ perceiving in addition to providing surround sensing capability [10].

5. AR causes stronger visual attention allocation during decision making phase [11] and attention is directed to roadway hazards [9].

6. AR provides additional information for decision making and helps in learning and preparation of decision makers. Spatial problem-solving may be increased and comprehensive decision making is facilitated [12].

7. AR has very effective real-time information communication with drivers [13] by providing engaging communication.

8. AR assists drivers to comply with rules and regulations by presenting safety-critical visual icons to the driver [13].

9. AR causes directing attention to important parts of user view, thus decreases cognitive load and causes decreased overload information processing [14].

7.3 Revisited Human Failure Taxonomies

In this section, we review the most used human failure taxonomies with the dependability-focused lens. More specifically, in compliance with Section 7.2.1, we use the term "failure" for human deviations from expected behaviors and not the term error, as it was done before the birth of the dependability community. We also distinguish failures from failure modes, by prefixing failure modes with "FM". Moreover, we use quotations when we cite the definitions and italics when we deemed necessary to complement the definitions with explanations taken from the Oxford dictionary (Simpson and Weiner 1989). Categories such as mistakes already mean failures. Thus, we do not repeat the word "failures".

7.3.1 Norman Taxonomy

Human failures based on [15] are:

1. **Mistakes** are failures in "formation of intention". Mistake meaning is *an act or judgment that is misguided or wrong*.

   (a) **Decision making mistakes** "arise when the situation is misclassified, or when inappropriate decisions and response selections are made".

   (b) **Description mistakes** are failures "in the retrieval and use of memory information". Description means *a spoken or written account of a person, object, or event*.

   (c) **System induced mistakes** are failures induced by the system that human is working within that. Induce means *succeed in persuading or leading (someone) to do something*.

2. **Slips** are failures in "performance of the intention". Slips are *pass or change to a lower, worse, or different condition, typically in a gradual or imperceptible way*.

7.3.2 Reason Taxonomy

Reason [16] divides human failures into three categories, which are:

1. **Slips and lapses** are "failures in either the execution or the storage stages of an action sequence." **Lapses** are *brief or temporary failures of concentration, memory, or judgment*. Slips and lapses are sub-divided into three categories including: recognition, memory and attention failures.
   (a) **Recognition failures** are failures in *identification of someone or something. Recognition means identification of someone or something or person from previous encounters or knowledge*. Recognition failures are divided into:
      i. **FM-Misidentifications** are wrong identifications of an object, message or signal. Misidentify means *identify (someone or something) incorrectly*. Identify means *recognize or distinguish (especially something considered worthy of attention)*.
      ii. **FM-Non-detections** are "failures to detect a signal or problem". Detect means *discover or identify the presence or existence of*.
      iii. **FM-Wrong detections** are "wrongly detecting problems or defects that were not actually present". Based on the definitions given in 7.2.2 [7] and based on the above recalled definitions, we can conclude that: misidentification is manifestation of a recognition failure as a value failure; non-detection is the manifestation as an omission failure; and, finally, wrong detection is the manifestation as a commission failure.
   (b) **Memory failures** are failures in "information processing stages including input, storage and retrieval".
      i. **Input failures** occur when "insufficient attention is given to the to-be-remembered material and it is lost from short-term memory." Input as a verb means *put (data) into a computer that here it is put into short-term memory*.
      ii. **Storage failures** occur when "the to-be remembered material decays or suffers interference in long-term memory". Forgetting intentions is a storage failure. Store means *keep (something) for future use*.

iii. **Retrieval failures** occur when "known material is not recalled at the required time". Retrieve means *get or bring (something) back from somewhere*.

(c) **Attention failures** are failures that occur "when attention is captured by something unrelated to the task in hand". Attention means *notice taken of someone or something*.

2. **Mistakes** are failures in "process of making plans". Mistakes can be rule-based or knowledge-based.

   (a) **Rule-based mistakes** are failures in "applying a problem-solving rule that is part of our stock of expertise".

   (b) **Knowledge-based mistakes** are failures in "finding a solution 'on the hoof'" and occur in novel situations that there is not any rule to solve the problem. 'On the hoof' means *without proper thought or preparation*.

3. **Violations** are "actions that involve some deliberate deviation from standard operating procedures". Violate means *breaking or failing to comply with (a rule or formal agreement). Violations can be routine or exceptional*.

   (a) **Routine violations** are when the users often do the violation as a habit and it is tolerated by authority. Routine means *a sequence of actions regularly followed.*

   (b) **Exceptional violations** are when the user violates but it is not his/her typical behavior pattern.

7.3.3 Rasmussen Taxonomy

Rasmussen et al. (Rasmussen 1982)'s human failure taxonomy stems from the analysis of mental processes, which consist of three levels of cognitive control behaviors:

- **Skill-based** refers to activities that are routine and humans do them automatically.

- **Rule-based** refers to activities that need identification and recall from memory.

- **Knowledge-based** refers to activities that are exploratory and unfamiliar.

[17]'s taxonomy includes:

1. **Detection failures**: "Operator does not respond to a demand".
2. **Identification of system state failures**: "Operator responds but misinterprets the system state."
3. **Decision failures**: Decision means *a conclusion or resolution reached after consideration.*
   (a) **Selection of goal failures**: "Operator responds to properly identified system state, but aims at wrong goal (e.g. operation continuity instead of safety)."
   (b) **Selection of system target state failures**: "Operator selects an improper system target state to pursue proper goal (e.g. he decreases power to 80% instead of shutdown)."
   (c) **Selection of task failures**: "The operator selects a task, an activity which will not bring the plant to the intended target state."
4. **Action failures**: Action means *the fact or process of doing something, typically to achieve an aim.*
   (a) **Procedure failures**: "The sequence of actions performed is inappropriate or incorrectly coordinated for the task chosen". Procedure means *an established or official way of doing something*.
   (b) **Execution failures**: "The physical activity related to the steps in the procedure is incorrect".
   (c) **Communication failures**: "Written or verbal messages are given incorrectly".

7.3.4 HFACS Taxonomy

The Human Factor Analysis and Classification System (HFACS) [18] taxonomy is based on Reason (Reason 2000) taxonomy. HFACS includes:

1. **Decision failures** occur when the intended action is performed intentionally but the plan is not appropriate for the situation. These failures can be divided into three categories:

   (a) **Procedural failures** also known as rule-based mistakes occur "during highly structured tasks of the sorts, if X, then do Y."

   (b) **Poor choices** (alias knowledge-based mistakes) occur during choosing the best action between multiple response options. It can happen because of lack of experience or time pressure.

   (c) **Problem solving failures** occur when there is a failure in understanding the problem or finding a procedure and response.

2. **Skill-based failures** are failures in "skills that occur without significant conscious thought". Skill-based actions are vulnerable to the following failures:

   (a) Attention failures

   (b) Memory failures

   (c) Technique failures or failures in "the manner in which one carries out a specific sequence of events"

3. **Perceptual failures** occur when "sensory input is degraded or unusual" for example because of visual illusions or misjudgment.

4. **Exceptional violations** are "isolated departures from authority."

5. **Routine violations** are habitual ignoring the rules and regulations often tolerated by governing authority.

7.3.5 SERA Taxonomy

SERA (Systematic Error and Risk Analysis) [19] represents Canadian forces' version of HFACS. SERA taxonomy includes:

1. **Intent failures** are failures in setting the goal that can be violation or non-violation.

   (a) **Violations** are setting a goal that is not consistent with rules and regulation. These can be routine or exceptional.

      i. **Routine violations** are "part of the individual's normal behavior. They are often tolerated and sanctioned by supervisory authority".

ii. **Exceptional violations** are "isolated departures from authority and not necessarily typical of an individual's behavior pattern. Usually management does not condone this behavior".

(b) **Non-violations** are setting a goal inconsistent with proficiency, capability or readiness of the individual/team.

2. **Attention failures** are failures "to attend to relevant information that was present or accessible".

3. **Sensory failures** are failures in physical capabilities for sensing the needed information. Knowledge (Perception) failures are when "the operator didn't have the pre-existing baseline knowledge or skills required to adequately or correctly interpret the situation."

4. **Perception failures** are when "All relevant sources of information were attended to but an incorrect perception was formed due to ambiguous or illusory information, or due to processing biases that shape our perceptions and filter the available information."

5. **Communication/Information failures** are failures "in communication or information exchange between machine (display) and human, or human and human."

6. **FM-Time Management** are failures "to use appropriate and effective time management strategies."

7. **Knowledge (Decision) failures** are when "the operator didn't have the pre-existing baseline knowledge or skills required to form an appropriate or correct response to the situation. These are failures in knowing what to do rather than failures in implementing the response."

8. **Ability to Respond Failures** are when "the operator does not have the physical capability to make the response required to perform the task."

9. **Action Selection Failures** are failures "in the decision process due to shortcomings in action selection, rather than misunderstanding or misperception of the situation. These are failures to formulate the right plan to achieve the goal, rather than a failure to carry out the plan."

10. **Slips, Lapses and Mode Errors** are "failures in action execution and when the responses are not implemented as intended."

(a) **Slips** are failures in skill-based behaviors.

(b) **Lapses** are failures in memory because of forgetfulness

(c) **Mode errors** are failures in actions that are appropriate in another mode but are inappropriate in the current mode and the operator forgets that.

11. **Feedback Failures** are failures "in backing-up, crosschecking or monitoring to ensure goal achievement."

7.3.6 Driving Taxonomy

Generic driver failure taxonomy [20] includes:

1. **Action failures** occur during executing the task and include: (a) **FM-Failing to act**, (b) **FM-Wrong action**, (c) **FM-Action mistimed**, (d) **FM-Action too much**, (e) **FM-Action too little**, (f) **FM-Action incomplete**, (g) **FM-Right action on wrong object**, (h) **FM-Inappropriate action**,
2. **Cognitive and decision making failures** are failures in recognizing the situation and taking decision and include: (a) **Perceptual failures**, (b) **FM-Wrong assumption**, (c) **Inattention**, (d) **FM-Distraction**, (e) **FM-Misjudgment**, (f) **Looking but failing to see**.
3. **Observation failures** are failures in observing a specific object or scene that include: (a) **FM-Failing to observe**, (b) **FM-Observation incomplete**, (c) **FM-Right observation on wrong object**, (d) **FM-Observation mistimed**.
4. **Information retrieval failures** are when there are failures in retrieving information from memory and include: (a) **FM-Misreading information**, (b) **FM-Misunderstanding information**, (c) **FM-Information retrieval incomplete**.
5. **Violations** are ignoring rules and regulations and include: (a) **Intentional violations** and (b) **Unintentional violations**.

7.4 Our Proposed Taxonomy

In this section, we try to harmonize and organize the existing taxonomies as a product line and propose a feature diagram, called AREXTax, for modeling

their commonalities and variabilities to present their evolution over time. For space reasons, our feature diagram is constituted of two sub-feature diagrams: one focusing on the human's functions and one on the failure modes potentially associated to these functions. In addition, we present an extension in order to deal with augmented reality.

7.4.1 Human Functions Taxonomy

Based on the six taxonomies, we retrieve and organize the human functions in Table 7.1. The rational for the fields of Table 7.1's columns is: 1) the function extracted from taxonomies; 2) subsection number of the related taxonomy and the failure that the function is extracted from; 3) failure modes (FM) of the function. For example, as it is explained in Subsection 7.3.2:1.a, recognition failure is a failure in the identification function, thus, in the first row of Table 7.1, identifying is extracted from the recognition failure. We also explained that misidentification is manifestation of a recognition failure as a value failure, so we add 7.3.2:1.a.i to the third column of first row. According to the definitions of the functions we define the hierarchy of them in the table. For example as mentioned in Subsection 7.3.2:1.a.ii detecting means identifying the presence, so we consider detecting as a subpart of identifying. Then, we extract human functions that are augmented via augmented reality. For example, when a driver uses visual augmented reality technology, he/she will detect more quickly through this technology and this AR-detection is an extended function of the extended-human.

According to subsection 7.2.3 we can extract human functions that are affected by using augmented reality. For example, in Subsection 7.2.3:2, it is stated that augmented reality augments driver perception to side lanes vehicles, so the affected human function in this case is perceiving. This function is shown in third row of Table 7.2. Then we present the human functions feature diagram in Figure 7.2 that shows functions deciding/ making plan, acting and executing are three common functions in all six taxonomies. It means that in Table 7.1 we have failures from all six taxonomies for these three functions or the functions that are subparts of them.

In addition, we extracted some more functions based on visual augmented reality application. These features are shown by dotted lines in Figure 7.2. For example, based on 7.2.3:4, we can consider GPS/lidar/infrared sensing as augmented functions, which transform humans into extended humans. By

Table 7.1: Human functions within failure taxonomies

Human function	Tax: failure	Human function
1.Identifying	7.3.2:1.a; 7.3.3:2	7.3.2:1.a.i; 7.3.5:6
1.1.Detecting	7.3.3:1	7.3.2:1.a.ii/1.a.iii; 7.3.5: 6
1.2.Sensing	7.3.5:3; 7.3.6:3	7.3.6:3.a-d; 7.3.5:6
2.Information processing	7.3.2:1.b; 7.3.4:2.b; 7.3.5:10.b	7.3.5:6
2.1.Inputting short-term memory	7.3.2:1.b.i	7.3.5:6
2.2.Storing in long-term memory	7.3.2:1.b.ii	7.3.5:6
2.3.Retrieving from memory	7.3.1:1.b; 7.3.2:1.b.iii; 7.3.6:4	7.3.6:4.c; 7.3.5:6
3.Paying attention	7.3.2:1.c; 7.3.4:2.a; 7.3.5:2; 7.3.6:2.c	7.3.5:6; 7.3.6:2.d
4.Deciding/ Making plan	7.3.1:1.a; 7.3.2:2; 7.3.3:3; 7.3.4:1; 7.3.5:10.c	7.3.5:6; 7.3.6:2.b/2.e
4.1.Applying a problem-solving rule	7.3.2:2.a; 7.3.4:1.a	7.3.5:6
4.2.Finding a solution 'on the hoof'	7.3.2:2.b; 7.3.4:1.c; 7.3.5:7	7.3.5:6
4.3.Selecting goal	7.3.1:1; 7.3.3:3.a; 7.3.5:1	7.3.5:6
4.4.Selecting target state	7.3.3:3.b	7.3.5:6
4.5.Selecting task	7.3.3:3.c; 7.3.4:1.b; 7.3.5:9	7.3.5:6
5.Conforming to rules	7.3.2:3; 7.3.4:4-5; 7.3.5:1.a; 7.3.6:5	7.3.5:6
6.Acting	7.3.3:4/4.a; 7.3.1:2	7.3.5:6
6.1.Executing	7.3.2:1; 7.3.3:4.b; 7.3.4:2/2.c; 7.3.5:10.a; 7.3.6:1; 7.3.5:8	7.3.6:1.a-h; 7.3.5:6
6.2.Communicating	7.3.1:1.c; 7.3.3:4.c; 7.3.5:5	7.3.5:6
6.3.Ensuring goal achievement by feedback	7.3.5:11	7.3.5:6

Table 7.2: Effects of AR on human functions

Function	Effects of AR	Extracted from
1.Detecting	Low visibility, accelerated, surround detecting	7.2.3:1/4
2.Sensing	GPS/lidar/infrared sensing, surround sensing	7.2.3:4
3.Perceiving	Accelerated, surround perceiving	7.2.3:2/3/4
4.Information processing	Decreased overload information processing	7.2.3:9
5.Paying attention	Directed paying attention	7.2.3:5
6.Problem solving/ Deciding	Comprehensive deciding	7.2.3:3/6
7.Communicating	Engaging communicating	7.2.3:7
8.Conforming to rules	Assisted conforming to rules	7.2.3:8

using AR information regarding surrounding the car and blind spots and displaying them on the view of driver [21], he/she can sense, detect and perceive these additional information. Thus, these functions are extended as surround detecting/sensing/perceiving.

7.4.2 Failure Modes Taxonomy

In this subsection, we show that [7] categorization is still valid and failure modes are still the same, shown in Figure 7.3. All FMs (failure modes in the third column of Table 7.1) are the features of the categories mentioned in Subsection 7.2.2. For example, according to definition mentioned in Subsection 7.3.2:1.a.i FM-misidentification is a wrong function and based on [7] wrong function can be considered as a feature for omission. All the features are optional in this feature diagram.

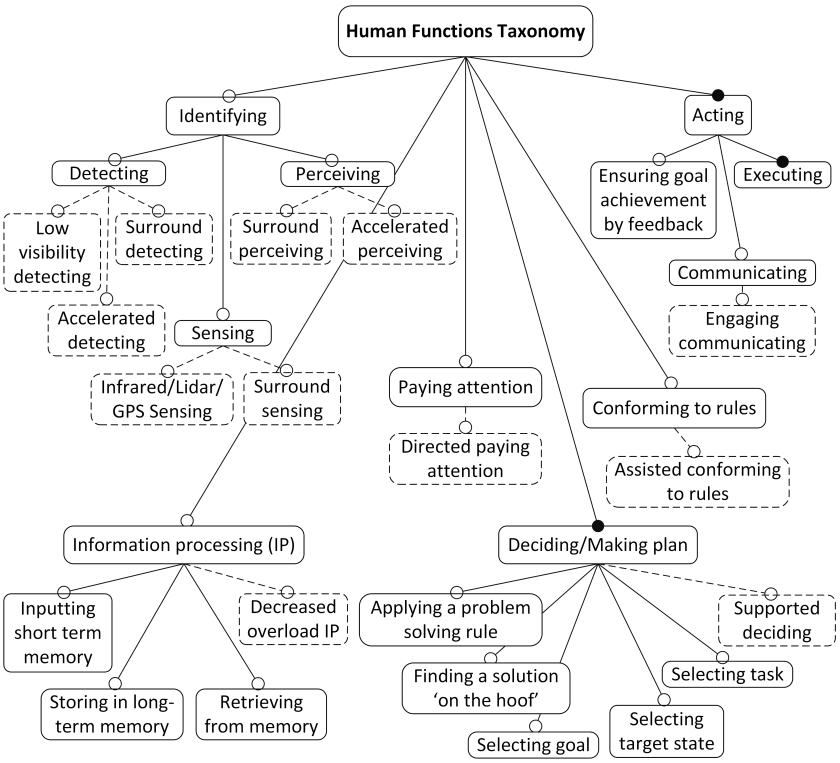

Figure 7.2: Human functions feature diagram

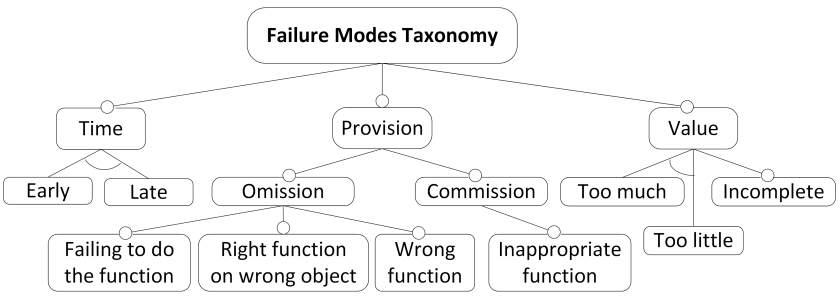

Figure 7.3: Failure modes feature diagram

7.5 Discussion

According to [22], there are a number of requirements that a good taxonomy should meet. In what follows, we discuss to which extent our taxonomy meets those requirements.

The proposed taxonomy is **accepted**, because it is structured and it is built on previous accepted taxonomies. It is **comprehensible**, because it is understandable by experts and those with interest in the field, since we split it based on human functions that are clearly defined. It is difficult to prove that the taxonomy is **complete**, but we can claim that it is complete to some extent because the covered taxonomies help to categorise the human failures based on human functions. It is **deterministic**, because we can determine human failures according to the related functions. However, sometimes, it is hard to discriminate if the failure is in detection or perception functions.

We cannot claim that it is **mutually exclusive** because each failure is not categorised into a single category. It is **repeatable** because we defined the procedure and by repeating the classification the result will be the same. In addition we used **terms complying with previous and state-of-the-art works** to remove/reduce the ambiguity. In some cases, in previous taxonomies, same terms were used with different meaning or same meaning with different terms. We reduced the ambiguity by using state-of-the-art-terms and showing how previously used terms were related with state-of-the-art terms. All the terms (including failures modes) are defined both according to the definitions mentioned in the related taxonomy and also according to Oxford dictionary. It is also **unambiguous** because the functions are clearly defined. Related to the **usefulness** of the suggested taxonomy, we do not have evidence yet. It should be evaluated by the community.

7.6 Conclusion

In this paper, we have reviewed the state-of-the-art on human failure taxonomies and provided a taxonomy of taxonomies, given as a feature diagram, to visually show their evolution in time. Then, we extended the taxonomy for visual augmented reality-extended humans.

As future work, with growing domain expertise, we aim at defining cross-cutting constraints to relate human functions with failure modes. In addition, we plan to use this taxonomy as the foundation of a failure logic-based analysis tool for socio-technical systems and validate it in industrial settings.

Acknowledgment

This research has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 764951.

Bibliography

[1] Van Krevelen, D., Poelman, R.: A survey of augmented reality technologies, applications and limitations. International journal of virtual reality **9**(2) (2010) 1–20

[2] Abdi, L., Abdallah, F.B., Meddeb, A.: In-vehicle augmented reality traffic information system: a new type of communication between driver and vehicle. Procedia Computer Science **73** (2015) 242–249

[3] Schwarz, F., Fastenmeier, W.: Augmented reality warnings in vehicles: Effects of modality and specificity on effectiveness. Accident Analysis & Prevention **101** (2017) 55–66

[4] Avizienis, A., Laprie, J.C., Randell, B., Landwehr, C.: Basic concepts and taxonomy of dependable and secure computing. IEEE transactions on dependable and secure computing **1**(1) (2004) 11–33

[5] Kang, K.C., Cohen, S.G., Hess, J.A., Novak, W.E., Peterson, A.S.: Feature-oriented domain analysis (foda) feasibility study. Technical report, Carnegie-Mellon Univ Pittsburgh Pa Software Engineering Inst (1990)

[6] Schobbens, P.Y., Heymans, P., Trigaux, J.C., Bontemps, Y.: Generic semantics of feature diagrams. Computer networks **51**(2) (2007) 456–479

[7] Pumfrey, D.J.: The principled design of computer system safety analyses. PhD thesis, University of York (1999)

[8] Fu, W.T., Gasper, J., Kim, S.W.: Effects of an in-car augmented reality system on improving safety of younger and older drivers. In: 2013 IEEE

International Symposium on Mixed and Augmented Reality (ISMAR), IEEE (2013) 59–66

[9] Schall Jr, M.C., Rusch, M.L., Lee, J.D., Dawson, J.D., Thomas, G., Aksan, N., Rizzo, M.: Augmented reality cues and elderly driver hazard perception. Human factors **55**(3) (2013) 643–658

[10] Phan, M.T.: Estimation of driver awareness of pedestrian for an augmented reality advanced driving assistance system. PhD thesis, Compiègne (2016)

[11] Eyraud, R., Zibetti, E., Baccino, T.: Allocation of visual attention while driving with simulated augmented reality. Transportation research part F: traffic psychology and behavior **32** (2015) 46–55

[12] Deshpande, A., Kim, I.: The effects of augmented reality on improving spatial problem solving for object assembly. Advanced Engineering Informatics **38** (2018) 760–775

[13] Farhat, I.: Examining the Effects of Augmented Reality Traffic Signs on Driver's Performance and Distraction. PhD thesis (2018)

[14] Hogg, J.L.: Cognitive design considerations for augmented reality. In: EEE International Conference on e-Learning, e-Business, Enterprise Information Systems, and e-Government, Las Vegas, NV. (2012)

[15] Norman, D.A.: Errors in human performance. Technical report, CALIFORNIA UNIV SAN DIEGO LA JOLLA CENTER FOR HUMAN INFORMATION PROCESSING (1980)

[16] Reason, J.: The Human Contribution. CRC Press (2016)

[17] Rasmussen, J.: Human errors. a taxonomy for describing human malfunction in industrial installations. Journal of occupational accidents **4**(2-4) (1982) 311–333

[18] Shappell, S.A., Wiegmann, D.A.: The human factors analysis and classification system–hfacs. (2000)

[19] Hendy, K.C.: A tool for human factors accident investigation, classification and risk management. Technical report, DEFENCE RESEARCH AND DEVELOPMENT TORONTO (CANADA) (2003)

[20] Stanton, N.A., Salmon, P.M.: Human error taxonomies applied to driving: A generic driver error taxonomy and its implications for intelligent transport systems. Safety Science **47**(2) (2009) 227–237

[21] Rickesh, T., Vignesh, B.N.: Augmented reality solution to the blind spot issue while driving vehicles. In: 2011 IEEE Recent Advances in Intelligent Computational Systems, IEEE (2011) 856–861

[22] Hansman, S.L.: A taxonomy of network and computer attack methodologies. (2003)

Chapter 8

Paper B: Effect of Augmented Reality on Faults Leading to Human Failures in Socio-technical Systems

Soheila Sheikh Bahaei, Barbara Gallina, Karin Laumann and Martin Rasmussen Skogstad.
In Proceedings of the 4th International Conference on System Reliability and Safety (ICSRS 2019a), Rome, Italy, November 2019.

Abstract

With the ultimate purpose of assessing risk within augmented reality-equipped socio-technical systems, in our previous work, we systematically organized and extended state-of-the-art taxonomies of human failures to include the failures related to the extended capabilities enabled by AR technologies. The result of our organization and extension was presented in form of a feature diagram. Current state-of-the-art taxonomies of faults leading to human failures do not consider augmented reality effects and the new types of faults leading to human failures. Thus, in this paper, we develop our previous work further and review state-of-the-art taxonomies of faults leading to human failures in order to: 1) organize them systematically, and 2) include the new faults, which might be due to AR. Coherently with what done previously, we use a feature diagram to represent the commonalities and variabilities of the different taxonomies and we introduce new features to represent the new AR-caused faults. Finally, an AR-equipped socio-technical system is presented and used to discuss about the usefulness of our taxonomy.

8.1 Introduction

Augmented reality (AR) technology, augments human capabilities such as hearing and observing to hear and observe more than others [1]. Visual augmented reality technology, augments human visual perception, by integrating digital content with the real world view of the user. For example, providing safety visual alerts on the windshield of the car through this technology can augment human visual perception to perceive risks and to drive safely. Using new technologies might introduce new types of dependability threats (specifically new faults and failures) that should be considered while analyzing risk. It is necessary here to clarify exactly what is meant by fault. If we consider a human as a component within a component-based system representing a socio-technical system, based on Avizienis et al. [2] terminology, a human failure is a deviation in human functioning from correct functioning (failure in the last subcomponent of human, which provide the output of human component) and the cause for the human failure is fault, which would be internal or external. Internal fault is failure in another subcomponent of human component and external fault is failure in another component, which its output is input for human component.

For example, an experiment on a virtual reality game, shows reduction of perception and balance of children immediately after the game [3]. Physical and mental states are influencing factors on human functions, thus they are subcomponents of human, which failure in these states might cause failure in human functioning or human failures. Another experiment on augmented reality guidance during manual tasks, shows decrement in user performance due to AR-based technical faults. The focal length of available head-mounted displays that were used for experiment are at least 2 meters and it is not appropriate for manual tasks that require high precision [4]. In this example AR-based technical fault is an external fault for human, which is coming from technical component. Augmented reality can also cause reduction in human depth of focus, reaction time and distance perception while driving, if not properly designed [5]. Design fault is an external fault for human component.

As it is shown in Figure 8.1, internal or external faults in socio-technical systems are the reasons for human failures and human failures may lead to risk, thus to analyze risk in socio-technical systems containing AR, it is required to consider effect of AR on human failures and faults leading to these human failures. Socio-technical systems are systems containing technical components, human and organization. External faults to human, may originate from technical components or organizational components and

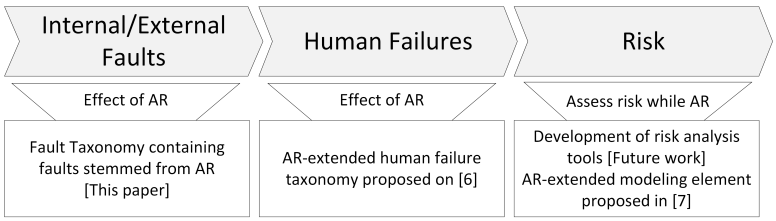

Figure 8.1: Risk-related causality chain in socio-technical systems

internal faults originate from other subcomponents of human. AR would influence on internal and external faults and would introduce new types of faults causing human failures.

In [6], we provided an AR-extended human failure taxonomy by considering state-of-the-art taxonomies as a product line and proposing a feature diagram containing human functions and including AR-extended human functions as extended features. In [7], we used this taxonomy for extending human modelling elements used in risk analysis tools.

Currently, there are different taxonomies of faults leading to human failures, which can be used as the foundation for risk analysis in safety-critical socio-technical systems. However, much uncertainty still exists about the effect of new technologies such as augmented reality and new types of faults to human failures that would be introduced to the system while using these technologies. In this paper, we concentrate on effect of AR on faults leading to human failures and by inheriting the strategy from [6], first, we review state-of-the-art taxonomies and vocabularies used for these faults with the lens of terminological framework on dependability. Then, we provide a feature diagram, because it is powerful to capture common and variable characteristics of different taxonomies. Finally, we extend the feature diagram by considering augmented reality effects and new faults that would cause human failures while using these technologies. The final outcome can be used as the foundation for risk analysis tools for safety-critical socio-technical systems.

The rest of the paper is organized as follows. In Section 8.2, we provide essential background information. In Section 8.3, we review state-of-the-art taxonomies of faults leading to human failures. In Section 8.4, we propose a taxonomy with the extension of faults stemmed from AR. In Section 8.5, we discuss about the use of this taxonomy on an automotive AR-equipped socio-

technical system. Finally, in Section 8.6, we present some concluding remarks and discuss about future work.

8.2 Background

In this section, we provide essential background information about visual augmented reality technology and feature diagram.

8.2.1 Visual Augmented Reality Technology

Visual augmented reality technology superimposes computational elements and objects on the real world view of the user. There are three types of AR displays containing head-worn, hand-held and spatial. Head-worn displays can be attached to the head, hand-held displays can be shown on a device or by a device that can be handled by hand and spatial displays are placed within the environment statically for cases with limited interactions. Head-up displays (HUDs) are an example of spatial displays that can be used by projecting information on the windshield of the car [1]. HUD is "any transparent display that presents data without requiring users to look away from their usual viewpoint" [8]. For example in Figure 8.2, navigation information is shown on the windshield of the car using AR.

Using an augmented reality warning in vehicles can improve driver awareness and reaction time efficiency, but can also increase cognitive-processing or distract the driver [9]. Schwarz and Fastenmeier [10] used augmented reality in a driver simulator study with 88 participants. The results show that visual warnings are advantageous and effective. Miller et al. [11] found that AR influences on interpersonal communications and decreases social presence, which might lead to human failure. Thus, while using new technologies, new types of faults leading to human failures would be added and should be considered in risk analysis.

8.2.2 Feature Diagram

A distinguishing characteristic of a family of systems that can be perceived by end-users is called a feature [12]. Families of products are also recognized as product lines [13]. To illustrate common and distinctive features of a product line, feature diagrams can be used. A simple example of a feature diagram is shown in Figure 8.3. As it is shown, feature diagrams are multi-level trees that

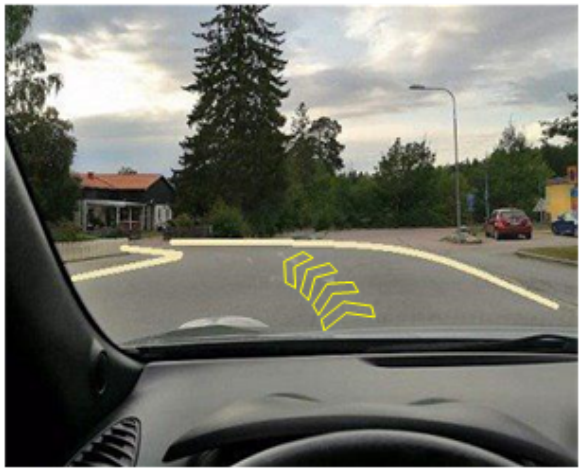

Figure 8.2: An example of AR information on head-up display

nodes are features and edges are for decomposition of features to more specific features. Features can have different types, for example mandatory, optional and alternative [12]. Mandatory features shown by solid dot are essential in the system and all the products in a product line have these features, but optional features (a node with a circle) are optional and some products may not have those features. Alternative features (XOR) are those features that only one of them are in each product of the product line. In the example shown in Figure 8.3, a family of AR devices are described that display is a mandatory feature, because all AR devices have display, but remote control and internet connection are optional, because there are some AR devices without remote control and internet connections. Display feature would be transparent or nontransparent that only one of them can happen, so these are alternative features.

8.3 Revisited Faults Taxonomies

In this section, we review state-of-the-art taxonomies of faults leading to human failures. In particular, we reconsider previously studied taxonomies such as: Rasmussen [14], HFACS (Human Factor Analysis and Classification System) [15], SERA (Systematic Error and Risk Analysis) [16] and Driving

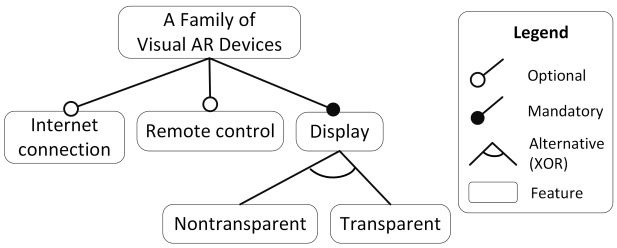

Figure 8.3: An example of a feature diagram for AR devices

[17] fault taxonomies. In addition, in this paper, we also consider SPAR-H (Standardized Plant Analysis Risk Human Reliability Analysis) [18] taxonomy, which provides influencing factors to human functions as a list of performance shaping factors.

As it was discussed in Section I, we use the term ”fault” based on Avizienis et al. [2] terminology, for defects in influencing factors leading to human failures. There may be some other faults leading to technical failures, but in this paper by fault, we mean faults leading to human failures. In order to follow the strategy in [6], for citing definitions we use quotations and to complement fault definitions, we use Oxford dictionary [19] meanings in cases which are necessary (shown in italic).

8.3.1 Rasmussen Faults Taxonomy

Rasmussen et al. [14] provided a taxonomy including faults to human failures in industrial installations based on analyzing mental processes. Faults based on this taxonomy (Figure 8.4) are divided to three groups including situation factors, performance shaping factors and causes of human malfunction faults.

1. **Situation factors faults** include:
   (a) **Task characteristics faults** arise when task is complicated or has some special characteristics that would cause human failure.
   (b) **Physical environment faults** arise when there are light, weather or other physical problems.
   (c) **Work time characteristics faults** arise when there is time pressure in doing the task.
2. **Performance shaping factors faults** include:

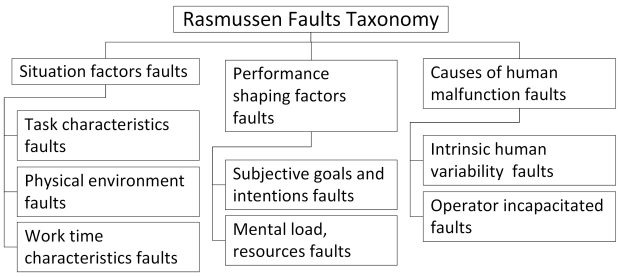

Figure 8.4: Rasmussen faults taxonomy

(a) **Subjective goals and intentions faults** arise when the goals and intentions are not defined correctly. Subjective means *based on or influenced by personal feelings, tastes, or opinions*.

(b) **Mental load, resources faults** arise when operator is not able to process huge amount of information mentally.

3. **Causes of human malfunction faults** include:

(a) **Intrinsic human variability faults**: intrinsic means *belonging naturally*. For example, low physical strength.

(b) **Operator incapacitated faults**: incapacitated means *deprived of strength or power*. For example, sickness.

8.3.2 HFACS Faults Taxonomy

HFACS [15] introduces another fault taxonomy (Figure 8.5), based on the avionic context, which is by analyzing over 300 aviation accidents. In this taxonomy, faults are divided into the following categories [20]:

1. **Faults in pre-conditions for unsafe acts** include:

(a) **Environmental factors faults** include:

i. **Physical environment faults** are faults related to physical environment such as unfavorable weather conditions.

ii. **Technological environment faults** are faults in technological environment such as problem in equipment.

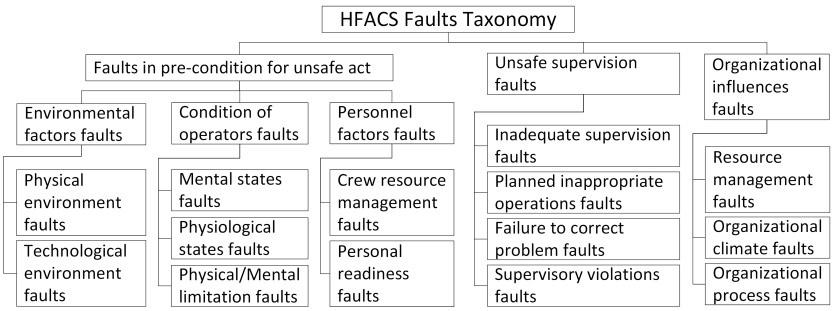

Figure 8.5: HFACS faults taxonomy

(b) **Condition of operators faults** include:

i. **Mental states faults** arise when the operator is not in a proper mental state.

ii. **Physiological states faults** arise when the operator is not in a proper physical state.

iii. **Physical/mental limitation faults** arise when the operator does not have a specific physical/mental capability.

(c) **Personnel factors faults** include:

i. **Crew resource management faults** arise when there is problem in managing human resource.

ii. **Personal readiness faults** arise when a person is not ready to act properly.

2. **Unsafe supervision faults** include:

(a) **Inadequate supervision faults** arise when supervisors do not provide their personnel, adequate guidance, training, leadership, oversight and whatever are needed to do safe and efficient job.

(b) **Planned inappropriate operations faults** arise when unsuitable operations are planned by supervisors.

(c) **Failure to correct problem faults** arise when safety deficiencies are known by supervisors but not corrected.

(d) **Supervisory violations faults** arise when supervisors disregard rules and regulations willfully.

3. **Organizational influences faults** include:

   (a) **Resource management faults** arise when there is problem in managing the resources such as personnel and monetary assets.

   (b) **Organizational climate faults** arise when working atmosphere such as organization's culture and policy cause human failure. Climate means *the prevailing trend of public opinion or of another aspect of life*.

   (c) **Organizational process faults** arise when there is problem in "corporate decisions and rules that govern the everyday activities within an organization".

8.3.3 SERA Faults Taxonomy

SERA [16] was developed as a tool for Canadian forces version of HFACS, but it can be used independent of HFACS. It divides faults to three categories including (Figure 8.6):

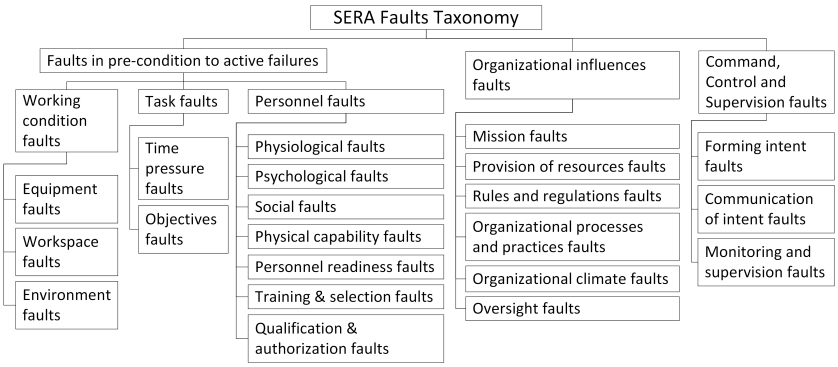

Figure 8.6: SERA faults taxonomy

1. **Faults in pre-conditions to active failures** include:

   (a) **Personnel faults** include:

      i. **Physiological faults** are not proper physiological state of the individual such as drowsiness, medical illness.

ii. **Psychological faults** are not proper psychological states, attitudes, traits, and processing biases.
iii. **Social faults** are problems in interaction among groups and teams.
iv. **Physical capability faults** are problems in physical abilities to sense and perform an action.
v. **Personnel readiness faults** are not being in a proper state in the sense of a physiological, psychological, physical and mental readiness to perform a task.
vi. **Training and selection faults** are lack of skills and knowledge required to do the job.
vii. **Qualification and authorization faults** are lack of legal prerequisites to perform a task.

(b) **Task faults** include:
i. **Time pressure faults** are lack of enough time to carry out the task.
ii. **Objectives faults** are unclear, inappropriate, inconsistent and risky task objectives.

(c) **Working condition faults** include:
i. **Equipment faults** are not proper condition of tools used to perform the task.
ii. **Workspace faults** arise when physical arrangement and layout of the workspace is not in a proper condition.
iii. **Environment faults** arise when conditions of the environment in which the activity is performed, is not suitable.

2. **C2S (Command, Control and Supervision) faults** include:

(a) **Forming intent faults** are problems in "goal setting process".
(b) **Communication of intent faults** are problems in "perceiving the intent by the subject audience".
(c) **Monitoring and supervision faults** are problems in "detecting and correcting ill-formed actions and disturbances".

3. **Organizational influences faults** include:

(a) **Mission faults** are problems in "what the organization is supposed to achieve". Mission means *a strongly felt aim, ambition, or calling*.

(b) **Provision of resources faults** are problems in "what the organization uses to achieve the mission". Provision means *the action of providing or supplying something for use*.

(c) **Rules and regulations faults** are problems in "Constraints on the process the organization uses to achieve the mission". Regulation means *a rule or directive made and maintained by an authority*.

(d) **Organizational processes and practices faults** are problems in "the way the organization should do it".

(e) **Organizational climate faults** are problems in "attitudes that affect how the people in the organization perceive the mission, what they actually do, and how they actually do it".

(f) **Oversight faults** are problems in "providing feedback so that managers can form a perception of organizational health and how well it is achieving its mission".

8.3.4 Driving Faults Taxonomy

Stanton and Salmon [17] present a taxonomy of faults leading to driving failures with an overview of the literature on human failures in road transport based on dominant psychological mechanisms involved, including perception, attention, situation assessment, planning and intention, memory and recall and action execution. Based on this taxonomy (Figure 8.7) faults include:

1. **Road infrastructure faults** include:

   (a) **Road layout faults** are problems in road surface.

   (b) **Road furniture faults** arise for example when traffic signs are not in proper condition.

   (c) **Road maintenance faults** arise when there is problem in renovating the road.

   (d) **Road traffic rules, policy and regulation faults** arise when there is not suitable rule, policy and regulation for road traffic.

2. **Vehicle faults** include:

   (a) **Human machine interface faults** are problems in interfaces such as navigation interface.

   (b) **Mechanical faults** are problems in mechanical part of vehicle such as problem in engine or gearbox.

(c) **Capability faults** are problems in power of vehicle such as limitation of engine horsepower.

(d) **Technology usage faults** arise when a technology is not used properly.

3. **Driver faults** include:

(a) **Physiological state faults** are problems in physiological state of driver such as sickness.

(b) **Mental state faults** are problems in mental state of driver such as tiredness.

(c) **Training and experience faults** are problems in training and experience of driver to carry out the task properly.

(d) **Knowledge, skills and attitudes faults** are lack of required knowledge, skills and attitudes.

(e) **Context faults** are problems in context. Context means *the circumstances that form the setting for an event, statement, or idea, and in terms of which it can be fully understood*. For example when driver is in hurry.

(f) **Non-compliance faults** are problems in complying with rule or standards. Compliance means *the state or fact of according with or meeting rules or standards*. For example, unqualified driving is a non-compliance fault.

4. **Other road user faults** include:

(a) **Other driver behavior faults** are problems caused by other drivers' unsafe acts.

(b) **Passenger influence faults** are problems caused by passengers.

(c) **Pedestrian behavior faults** are problems caused by pedestrian.

(d) **Law enforcement faults** are problems in complying with law. Enforcement means *the act of compelling observance of or compliance with a law, rule, or obligation*.

(e) **Other road user behavior faults** are problems caused by other road users' behavior.

5. **Environmental conditions faults** include:

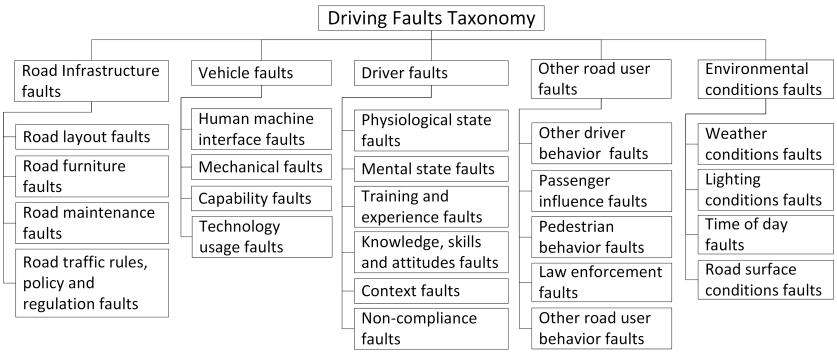

Figure 8.7: Driving faults taxonomy

(a) **Weather conditions faults** are not suitable weather condition such as fogginess.

(b) **Lighting conditions faults** are not suitable lighting condition such as darkness.

(c) **Time of day faults** are problems caused by time of day.

(d) **Road surface conditions faults** are inappropriate road surface conditions.

8.3.5 SPAR-H Faults Taxonomy

SPAR-H [18] is a human reliability analysis method used in commercial US nuclear power plants. Faults based on this method are categorized to eight faults including (Figure 8.8):

1. **Available time faults** refer to faults in the amount of time available relative to the time required.

2. **Stress faults** refer to faults in the level of undesirable conditions and situations that prevent the operator form completing a task.

3. **Complexity faults** refer to faults in difficulty of a task in a special context.

4. **Experience/Training faults** refer to faults in experience/training of the operators for carrying out the tasks.

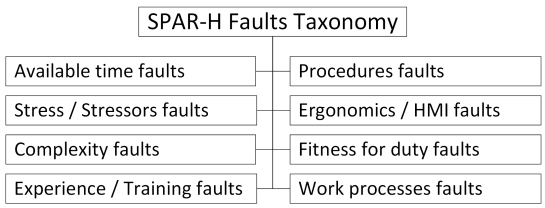

Figure 8.8: SPAR-H faults taxonomy

5. **Procedures faults** refer to faults in availability and using of formal procedures for operating a task.
6. **Ergonomic faults** refer to faults in the equipment, displays and controls, layout, quality, and quantity of information available from instrumentation, and the interaction of the operator/crew with the equipment to perform the task.
7. **Fitness for duty faults** refer to faults in suitability of the operator for doing the task physically and mentally.
8. **Work process faults** refer to faults in inter-organizational safety culture, work planning, communication, and management support and policies.

8.4 Our Proposed Fault Taxonomy

In this section, we propose a taxonomy of taxonomies by considering augmented reality effects. First, we adjust and organize the existing taxonomies as a product line. Then, we consider effects of augmented reality on these factors based on available experiments and studies. Finally, we propose a feature diagram for modelling existing taxonomies' commonalities and variabilities and effects of augmented reality.

8.4.1 Fault Categorization Based on State-of-the-art Taxonomies

Based on five taxonomies, we retrieve and organize fault categories in Table 8.1 and 8.2. The rational for the fields of table's columns is: 1) fault category based on state-of-the-art taxonomies; 2) subsection number of the related

taxonomy and related fault category. For example, physical state fault that is a personnel fault category is mentioned in five taxonomies. Different terms may be used for fault categories in various taxonomies, but based on explanation, we organized them to have a categorization based on all five taxonomies. In HFACS (Subsection 8.3.2), fault category 1.b.ii, which is physiological states faults and in SERA (Subsection 8.3.3), fault category 1.a.i refers to physiological faults.

8.4.2 Effect of Augmented Reality

In this section, we explain about effect of augmented reality on fault categories based on available studies and experiments. For some of the categories, there is not any study or experiment to show the effect of augmented reality. Thus, we cannot provide any extension related to those categories for our taxonomy. AR effects on faults would be positive or negative and in both cases we need to consider them, because even positive effects can introduce new types of faults to the system, in case of failing to provide the expected effects.

Task Faults

An experiment presented in [4], was designed to investigate user performance during AR-guided manual tasks and results indicate decrement in users' performance. In this experiment, optical see-through (OST) head-mounted display (HMDs) is used for connect-the-dots task, which is a manual task with high precision. As it is explained in this study, the reason for decrement of performance is that focal length in available OST HMDs is at least 2 meters and users can not focus on both virtual and real content for manual tasks. However these results are for HMDs and for manual tasks with high precision, from this example we can elicit AR-guided task as an influencing factor on human function. In this example, it is not the task itself that would cause human failure, instead it is fault in AR-guided task that can cause human failure.

Physical and Mental State Faults

There are some jobs with difficult situations and repetitive tasks that threaten operators' mental and physical healthy states. For example, mental and

Table 8.1: Fault categorization based on state-of-the-art fault taxonomies

Fault category	Taxonomy: fault
1.Personnel faults	8.3.2:1.c/1.c.ii, 8.3.3:1.a/1.a.v
1.1.Physical state faults	8.3.1:3, 8.3.2:1.b.ii, 8.3.3:1.a.i, 8.3.4:3.a, 8.3.5:2
1.2.Mental state faults	8.3.1:3, 8.3.2:1.b.i, 8.3.3:1.a.ii, 8.3.4:3.b, 8.3.5:2
1.3.Physical/ Mental capability faults	8.3.1:2.b/3, 8.3.2:1.b.iii, 8.3.3:1.a.iv, 8.3.4:3.d, 8.3.5:7
1.3.1.Training/ Experience faults	8.3.3:1.a.vi, 8.3.4:3.c/3.d , 8.3.5:4
1.4.Social faults	8.3.2:1.c.i, 8.3.3:1.a.iii, 8.3.4:4, 8.3.5:2
1.5.Authorization faults	8.3.3:1.a.vii, 8.3.4:3.f/4.d, 8.3.5:8
2.Task faults	8.3.3:1.b, 8.3.1:1.a
2.1.Time pressure faults	8.3.1:1.c, 8.3.3:1.b.i, 8.3.5:1
2.2.Objectives faults	8.3.1:2.a, 8.3.3:1.b.ii,
2.3.Complexity faults	8.3.5:3
2.4.Procedure faults	8.3.5:5
3.Environment faults	8.3.3:1.c
3.1.Equipment faults	8.3.2:1.c, 8.3.3:1.c.i, 8.3.4:2, 8.3.5:6
3.2.Condition faults	8.3.1:1.b, 8.3.2:1.a.i, 8.3.3:1.c.ii-iii, 8.3.4:1/3.e/5, 8.3.5:6
4.Organization and regulation faults	8.3.2:3, 8.3.3:2/3
4.1.Resource management faults	8.3.2:3.a, 8.3.3:3.b
4.2.Organizational climate faults	8.3.2:3.b, 8.3.3:3.e
4.3.Organizational process faults	8.3.2:3.c, 8.3.3:3.d, 8.3.5:8
4.4.Supervision faults	8.3.2:2.a-c, 8.3.3:2.c

Table 8.2: Fault categorization based on state-of-the-art fault taxonomies (Cont.)

Fault category	Taxonomy: fault
4.4.1.Forming intent faults	8.3.3:2.a
4.4.2.Communication of intent faults	8.3.3:2.b
4.4.3.Monitoring and supervision faults	8.3.3:2.c
4.4.4.Supervisory violation faults	8.3.2:2.d
4.5.Rules and regulations faults	8.3.3:3.c, 8.3.4:1.d
4.6.Oversight faults	8.3.3:3.f
4.7.Mission faults	8.3.3:3.a

physical states of astronaut crews in long-duration missions on the moon would be deteriorated and new technologies such as immersive virtual reality and augmented reality are examined to be used in order to upkeep mental and physical health [21]. AR/VR technologies also have been used for treatment of several mental disorders on clinical and health psychology and have provided important contributions to mental health [22]. These technologies can be considered as restorative mental and physical health measures and if not provided can cause human failure, so as AR-caused faults, restorative mental health measure faults and restorative physical health measure faults can be considered.

Social Faults

In an experiment presented in [11], with the aim of investigating AR effects on interpersonal communications, results show that people using AR have lower social presence and they feel significantly less connected. However this experiment was done by headset, the results can be used for other applications that operator is the person who sees AR and other people are not aware of this AR information and it disrupts common ground between interactants. As an augmented reality factor, social presence can be considered, because using AR can influence on social presence and by decreasing that it would cause

human failure. Interpersonal attraction that refers to how much participants like each other, also was studied when using AR and results show that there was no significant difference on interpersonal attraction while using AR.

Mental/Physical Capability Faults

Based on an experiment presented in [23], neurological effects of AR or effects of AR on the brain was measured, using brain-imaging technology. Results show that AR doubles brain visual attention in comparison to non-AR tasks and increases brain cognitive activity. Memory encoding is 70% higher when using AR, which means that AR delivers information in a powerful way to be retained in memory. AR elicits an astonishing response in the brain, so brain elicitation can be considered as a factor that is correlated with AR and brain elicitation faults can be added to our taxonomy as AR-related faults.

Training/Experience Faults

AR has the power to provide interactive ways to engage learners and strengthen their motivation for learning and to enhance their experience through computer graphics elements [24]. Interactive training/experience can be considered as a factor that can effect human performance and interactive training/experience faults are AR-related faults.

Environment Faults

Augmented reality technology integrates elements from virtual reality with elements from real world, thus we have an augmented environment that can be considered in our taxonomy. This augmented environment includes virtual objects that can be stationary or manipulated by user [25] and faults in augmented environment would cause human failure.

Organization and Regulation Faults

A study in [26] investigates key factors that facilitate adoption of AR technologies by e-commerce firms. This research shows that by emergence of AR, adoption of AR will be added as a new factor in organization and regulation that problem in this adoption would introduce a new fault that is organization and regulation AR adoption fault.

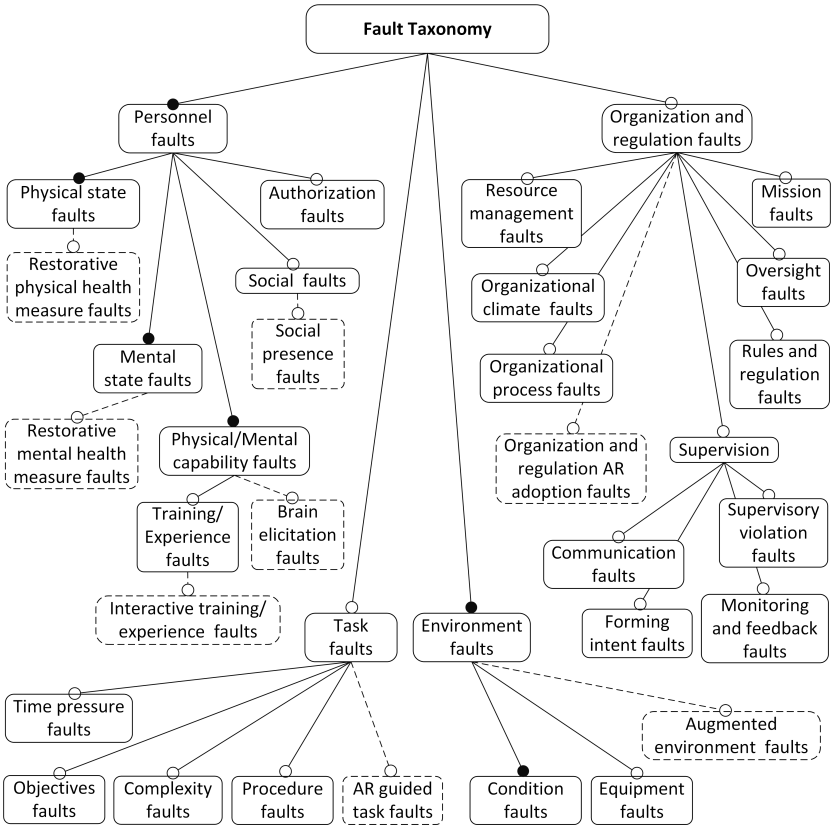

Figure 8.9: Proposed feature diagram for fault taxonomy

8.4.3 Proposed Feature Diagram

A feature diagram is presented in Figure 8.9, for modelling existing taxonomies' commonalities and variabilities and effects of augmented reality, which is called AREFTax. It shows physical/mental state, physical/mental capability, environment and condition faults are common faults in all taxonomies. It means that these faults or subcategories of them in Table 8.1 are in all taxonomies. There are also some more features based on augmented reality effects that are explained in Subsection IV.B. These features are shown by dotted lines. For example, as it was explained, augmented reality would decrease social presence, thus social presence faults can be considered as a new fault, which can cause human failures.

8.5 Automotive AR-equipped System

In this section, we use an augmented reality Collision Warning System (CWS) in a car that is an automotive domain-related socio-technical safety-critical system. Collision warning systems are special types of Advanced Driver Assistance Systems (ADAS), which provide notifications for drivers about potential hazards around the vehicle to avoid collisions. There are different technologies for presenting collision warning information such as AR, which is useful for providing visual clues and annotations on the user's view [27].

New technologies such as HUDs are AR capable and provide the opportunity to show AR information on the windshield of the car. The advantage of using this technology is that driver does not need to refocus to see outside, after looking through AR information [8].

We consider a HUD on the windshield of a car to provide notification or navigation information for driver to avoid collision. For example, when another car is in close distance, navigation information for changing the lane would be proposed on the windshield. We discuss about different possible faults to see if the proposed taxonomy deals with all possible faults for this example.

In this example, which is a socio-technical system, there are three components including human, organization and technical component (AR-technical component). We use the proposed fault taxonomy to model these components and their subcomponents and to show the possibility of modelling AR-related subsystem failure behavior. Organizational factors influencing human functioning are organization and regulation AR adoption and rules and regulation. Thus, organization and regulation AR adoption and rules and regulation can be considered as subcomponents of organization

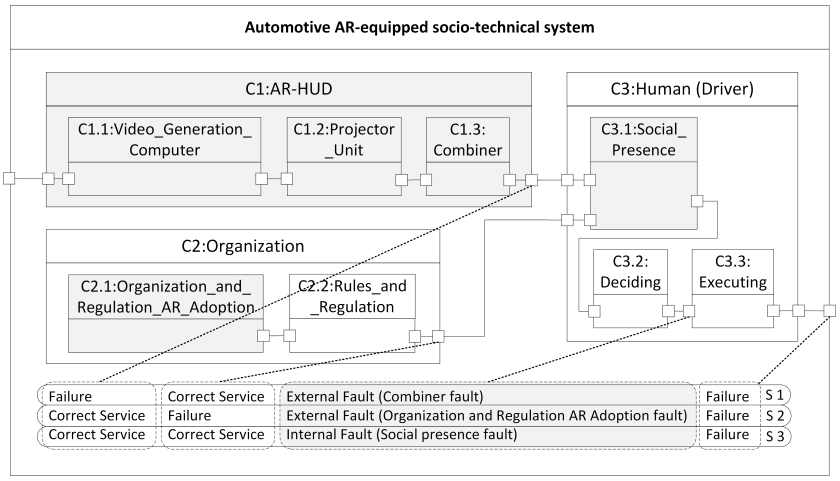

Figure 8.10: Using the proposed fault taxonomy in an automotive AR-equipped socio-technical system modelling

component. Human can be modeled using various states and functions. In this example, we consider social presence, deciding and executing functions as three subcomponents of human component. An AR-HUD component contains three primary subcomponents: a projector unit that produces an image on a combiner, a combiner that is a flat piece of glass and can be the windshield of the car and a video generation computer that generates the information that should be displayed by projector unit [8]. To illustrate the case study, we explain about three scenarios depicted in Figure 8.10. AR-related modelling elements and faults are shown by gray color, to show the effect of AR and the contribution of the proposed taxonomy.

In the first scenario, content provided by AR-HUD is wrong and leads to the driver's failure. For example, there is failure in combiner of AR-HUD, which is an AR-technical component. This failure is an external fault for human component and would cause human failure. In our taxonomy these kinds of failures that are dependent on the specific AR-technology and AR-device, are presented as augmented environment fault, which is a feature of environment in the proposed feature diagram.

In the second scenario, content provided by AR-HUD is correct, but there is failure in organization and regulation AR adoption, which is an external fault for human component and we represent it as organization and regulation AR

adoption fault as a feature of organization and regulation fault in the proposed feature diagram. This subcomponent is based on the AR-extended faults part of the taxonomy. For example, when the organization does not provide facilities for using AR in organization and when the organization does not provide AR-related rules and regulation.

In the third scenario, there is failure in social presence of the driver, which is an internal fault for deciding and then for executing subcomponents and causes human failure. For example, when driver miss the common ground with other people, this failure would lead to wrong decision and wrong action. This subcomponent is also AR-related fault and thanks to the proposed fault taxonomy, it is possible to use it for extending modelling elements and for considering AR-related faults while doing risk analysis.

As it is shown in this example, the proposed taxonomy can be used for enhancing modelling of internal and external faults leading to human failures, used in modelling techniques of risk analysis tools.

8.6 Conclusion

In this paper, first, we presented a review of the state-of-the-art taxonomies of faults leading to human failures, then we proposed an arrangement of taxonomies through a feature diagram that clarifies commonalities and variations between different taxonomies in a perceivable manner. Finally, the taxonomy is extended for augmented reality applications by adding faults stemmed from AR as new features to the proposed feature diagram. Application of this taxonomy is in risk analysis to increase safety in systems containing AR.

There are some opportunities to be considered as future work. In the future, this taxonomy can be used for extending modelling elements of influencing factors on human failures in SafeConcert [28], which is a metamodel for modelling technical and socio entities in socio-technical systems. Extended modelling elements can be used as the foundation of risk analysis tools such as Concerto-FLA [29], which is a risk analysis tool for socio-technical systems.

Acknowledgment

This work is funded by EU H2020 MSC-ITN grant agreement No 764951.

Bibliography

[1] Van Krevelen, D., Poelman, R.: A survey of augmented reality technologies, applications and limitations. The International Journal of Virtual Reality **9**(2) (2010) 1–20

[2] Avizienis, A., Laprie, J.C., Randell, B., Landwehr, C.: Basic concepts and taxonomy of dependable and secure computing. IEEE transactions on dependable and secure computing **1**(1) (2004) 11–33

[3] McKie, R.: Virtual reality headsets could put children's health at risk. https://www.theguardian.com/technology/2017/oct/28/virtual-reality-headset-children cognitive-problems (2017)

[4] Condino, S., Carbone, M., Piazza, R., Ferrari, M., Ferrari, V.: Perceptual limits of optical see-through visors for augmented reality guidance of manual tasks. IEEE transactions on bio-medical engineering (2019)

[5] Sabelman, E.E., Lam, R.: The real-life dangers of augmented reality. IEEE Spectrum **52**(7) (2015) 48–53

[6] Sheikh Bahaei, S., Gallina, B.: Augmented reality-extended humans: towards a taxonomy of failures – focus on visual technologies. In: European Safety and Reliability Conference (ESREL), Research Publishing, Singapore (2019)

[7] Sheikh Bahaei, S., Gallina, B.: Towards Assessing Risk of Reality Augmented Safety-critical Socio-technical Systems. Published as proceedings annex on the International Symposium on Model-Based Safety and Assessment (IMBSA) website. http://easyconferences.eu/imbsa2019/proceedings-annex/ (2019)

[8] Phan, M.T.: Estimation of driver awareness of pedestrian for an augmented reality advanced driving assistance system. PhD thesis, Université de Technologie de Compiègne (2016)

[9] Hall, N., Lowe, C., Hirsch, R.: Human factors considerations for the application of augmented reality in an operational railway environment. Procedia Manufacturing **3** (2015) 799–806

[10] Schwarz, F., Fastenmeier, W.: Augmented reality warnings in vehicles: Effects of modality and specificity on effectiveness. Accident Analysis & Prevention **101** (2017) 55–66

[11] Miller, M.R., Jun, H., Herrera, F., Villa, J.Y., Welch, G., Bailenson, J.N.: Social interaction in augmented reality. PloS one **14**(5) (2019) e0216290

[12] Kang, K.C., Cohen, S.G., Hess, J.A., Novak, W.E., Peterson, A.S.: Feature-oriented domain analysis (FODA) feasibility study. Technical report, Carnegie-Mellon Univ Pittsburgh Pa Software Engineering Inst (1990)

[13] Schobbens, P.Y., Heymans, P., Trigaux, J.C., Bontemps, Y.: Generic semantics of feature diagrams. Computer Networks **51**(2) (2007) 456–479

[14] Rasmussen, J.: Human errors. a taxonomy for describing human malfunction in industrial installations. Journal of occupational accidents **4**(2-4) (1982) 311–333

[15] Shappell, S.A., Wiegmann, D.A.: The human factors analysis and classification system–HFACS. Technical report, Civil Aeromedical Institute (2000)

[16] Hendy, K.C.: A tool for human factors accident investigation, classification and risk management. Technical report, Defence Research And Development Toronto (Canada) (2003)

[17] Stanton, N.A., Salmon, P.M.: Human error taxonomies applied to driving: A generic driver error taxonomy and its implications for intelligent transport systems. Safety Science **47**(2) (2009) 227–237

[18] Gertman, D., Blackman, H., Marble, J., Byers, J., Smith, C., et al.: The SPAR-H human reliability analysis method. US Nuclear Regulatory Commission **230** (2005)

[19] Simpson, J.A.: The Oxford english dictionary. Volume 15. Oxford University Press, USA (1989)

[20] Wiegmann, D.A., Shappell, S.A.: A human error approach to aviation accident analysis: The human factors analysis and classification system. Routledge (2017)

[21] Salamon, N., Grimm, J.M., Horack, J.M., Newton, E.K.: Application of virtual reality for crew mental health in extended-duration space missions. Acta Astronautica **146** (2018) 117–122

[22] Ventura, S., Baños, R.M., Botella, C.: Virtual and augmented reality: New frontiers for clinical psychology. State of the Art Virtual Reality and Augmented Reality Knowhow (N Mohamudally, Eds.) Rijeka: InTech (2018) 99–118

[23] Heather, A.: How augmented reality affects the brain. Technical report, Neuro-Insight (2018)

[24] Lee, K.: Augmented reality in education and training. TechTrends **56**(2) (2012) 13–21

[25] Gutiérrez, M., et al.: Augmented reality environments in learning, communicational and professional contexts in higher education. Digital Education Review **26** (2014) 22–35

[26] Chandra, S., Kumar, K.N.: Exploring factors influencing organizational adoption of augmented reality in e-commerce: Empirical analysis using technology-organization-environment model. Journal of Electronic Commerce Research **19**(3) (2018)

[27] Park, B.J., Lee, J.W., Yoon, C., Kim, K.H.: Augmented reality for collision warning and path guide in a vehicle. In: Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology, ACM (2015) 195–195

[28] Montecchi, L., Gallina, B.: SafeConcert: A metamodel for a concerted safety modeling of socio-technical systems. In: International Symposium on Model-Based Safety and Assessment, Springer (2017) 129–144

[29] Gallina, B., Sefer, E., Refsdal, A.: Towards safety risk assessment of socio-technical systems via failure logic analysis. In: 2014

IEEE International Symposium on Software Reliability Engineering Workshops, IEEE (2014) 287–292

Chapter 9

Paper C: Extending SafeConcert for Modelling Augmented Reality-equipped Socio-technical Systems

Soheila Sheikh Bahaei and Barbara Gallina.
In Proceedings of the 4th International Conference on System Reliability and Safety (ICSRS 2019b), Rome, Italy, November 2019.

Abstract

With the emergence of new technologies such as augmented reality in socio-technical systems, traditional risk assessment methods may fail to have a comprehensive system modeling, because these technologies extend human's capabilities, which might introduce new types of human failures caused by failing these extended capabilities and new types of faults leading to human failures. Current state-of-the-art modeling techniques do not contemplate these capabilities and augmented reality-caused faults leading to human failures. In our previous work, we proposed an extension for modeling safety-critical socio-technical systems, to model augmented reality-extended humans by using a taxonomy that contains AR-specific human's failure behavior. In this paper, we continue our extension by investigating faults leading to human failures including faults because of augmented reality. Our extension builds on top of a metamodel for modeling socio-technical component-based systems, named SafeConcert. We illustrate our extension on two fictitious but credible systems taken from air traffic control and rail industry. In order to model augmented reality-equipped socio-technical systems, we need to consider human and organization as parts of the system and augmented reality as a technology used in the system.

9.1 Introduction

Augmented reality (AR) can enhance humans' capabilities to see, hear, probably touch, smell and taste more than other humans [1], thus it provides the possibility to have AR-extended humans. In visual augmented reality, computer-generated suitable pieces of information are superimposed on the real world view of the user [2]. For example, using navigation metaphors of landmarks and routes in augmented reality mobile systems can improve human wayfinding and human will have an extended capability that was absent before using this technology [3]. Based on the experiment conducted in [4], context-adaptive augmented reality enhances air traffic controllers' performance and provides new opportunities for air traffic management [4].

New technologies improve human performance, meanwhile they might introduce new failures and faults to socio-technical systems, which should be considered during risk assessment. Throughout this paper, we consider a human as a component in a component-based architecture. Based on Avizienis et al. [5] terminology, human failure is deviation in human functioning and fault is the cause of the human failure. Failure might act as fault in a subsequent component. Faults leading to human failures can be external, if they emanate from subcomponents of other components, or internal, if they emanate from other subcomponents of human component itself. An experiment conducted in [6], shows that presence of augmented cueing aid for expected target detection on the display may distract the viewer from the presence of unexpected targets in the environment and leads to human failures. Another experiment conducted in [7], indicates that augmented reality information on the head up display (HUD) with less than 8 deg (angular degree) from information in the real world would cause cognitive tunneling. Cognitive tunneling means locked attention on AR information and neglecting the real world, which would lead to human failure.

To do risk assessment in the presence of augmented reality, the first step is identification of what can go wrong while there is augmented reality and how it would effect on modeling techniques used for risk analysis. Currently, there are different modeling languages for safety-critical socio-technical systems. However, there is no detailed investigation of the effect of new technologies such as augmented reality. Consequently, the concept is not considered in modeling of new faults leading to human failures that would be introduced to the system while using these technologies. Human failures and faults classifications in SafeConcert metamodel are based on SERA (Systematic Error and Risk Analysis) [8] taxonomy. In [9], we proposed AREXTax, which

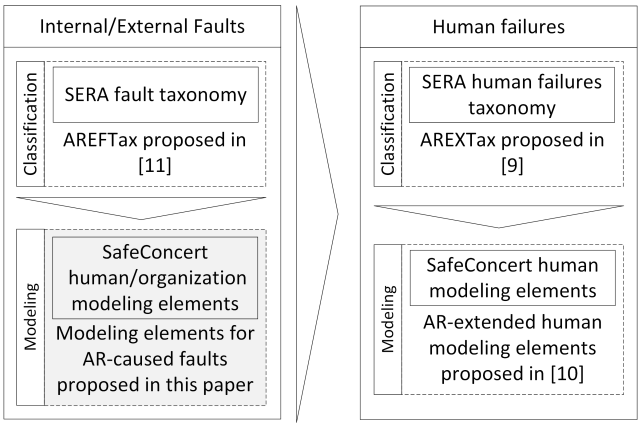

Figure 9.1: Contribution of this paper

is an AR-extended human function taxonomy by considering state-of-the-art failure taxonomies and AR-extended human functions. In [10], we incorporated this taxonomy while performing a first step towards a substantial extension of SafeConcert human modeling elements used in risk analysis tools, to enable modeling of AR-extended human capabilities. In [11], we proposed AREFTax, a taxonomy of faults leading to human failures, which contains AR-caused faults leading to human failures. In this paper, we extend SafeConcert metamodel to provide the ability of modeling faults leading to human failures including AR-caused faults to empower this metamodel for risk assessment in AR-equipped socio-technical systems. Our extension consists of extension in human modeling elements and extension in organization modeling elements. To clarify the contribution of this paper, the extension made in this paper is shown by gray color in Figure 9.1. In addition, we show our extension on two fictitious but credible systems within aerodrome control environment and train driving context.

The rest of the paper is organized as follows. In Section 9.2, we provide essential background information. In Section 9.3, we propose our extension on SafeConcert metamodel, based on a taxonomy of faults leading to human failures including AR-caused faults. In Section 9.4, we model two AR-equipped socio-technical systems from air traffic control and rail industry using the extended metamodel. In Section 9.5, we discuss about the strengths and limitations of the proposed extension. In Section 9.6, we provide related

works. Finally, in Section 9.7, we present some concluding remarks and discuss about future work.

9.2 Background

In this section, we provide the background information about AREXTax on augmented reality-extended humans, AREFTax on faults leading to human failures, SafeConcert and its implementation and extended SafeConcert.

9.2.1 AREXTax on Augmented Reality-extended Humans

In [9], we proposed a taxonomy of human functions based on state-of-the-art human failure taxonomies and by considering effect of augmented reality and we called it AREXTax. Our taxonomy, presented as a feature diagram, synthesizes the historical evolution of the previously proposed taxonomies (Norman [12], Reason [13], Rasmussen [14], HFACS (Human Factor Analysis and Classification System) [15], SERA [8] and Driving [16] human failure taxonomies) and it also considers AR-specific characteristics. More specifically, we extended the taxonomy for socio-technical systems containing visual augmented reality-extended humans.

Based on this taxonomy, we have a list of human functions that is extracted from the above-listed failure taxonomies. For example *paying attention* function is extracted from *attention failure*. The list of human functions is shown in Figure 9.2 and functions characterizing AR-extended humans are shown by dotted border, which are extracted from studies and experiments on augmented reality. For example, if AR information is shown on the windshield of a car, it helps the driver to detect the presence of a person in blind spots [17]. Thus, *surround detecting* can be considered as an extended function. These functions can be considered as subcomponents within the composite component representing the human in socio-technical systems.

9.2.2 AREFTax on Faults Leading to Human Failures

In [11], we proposed a taxonomy of faults leading to human failures based on state-of-the-art taxonomies. In addition to the faults extracted from previous taxonomies including Rasmussen [14], HFACS [15], SERA [8], Driving [16] and SPAR-H (Standardized Plant Analysis Risk Human Reliability Analysis) [18], we added faults stemmed from AR, based on related studies and

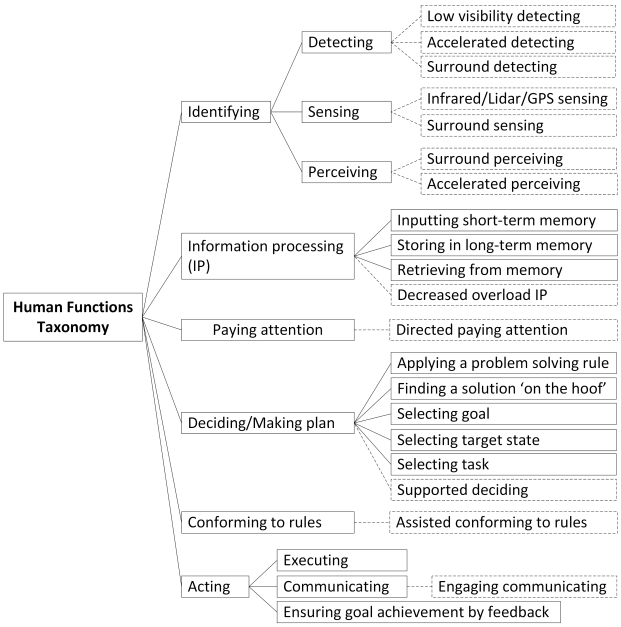

Figure 9.2: Function classification of AR-extended humans (adapted from [9])

experiments to have a comprehensive taxonomy useful for AR-equipped socio-technical systems. This taxonomy, which is called AREFTax, is shown in Figure 9.3. For example, social faults (problems in communicating with others) are personnel faults, categorized based on state-of-art taxonomies, which might lead to human failures. Using augmented reality may decrease social presence and a new type of fault (social presence fault) may lead to human failures [19]. Thus, we added this new type of fault as AR-caused fault (shown by dotted border) to the taxonomy.

9.2.3 SafeConcert and Its Implementation

SafeConcert [20] is a metamodel that facilitates unified analysis of interdependencies between socio and technical entities, because it offers constructs for modeling both of them in a common model. This metamodel is a subset of CHESS ML (CHESS Modeling Language) [21], which is a UML (Unified Modeling Language)-based modeling language.

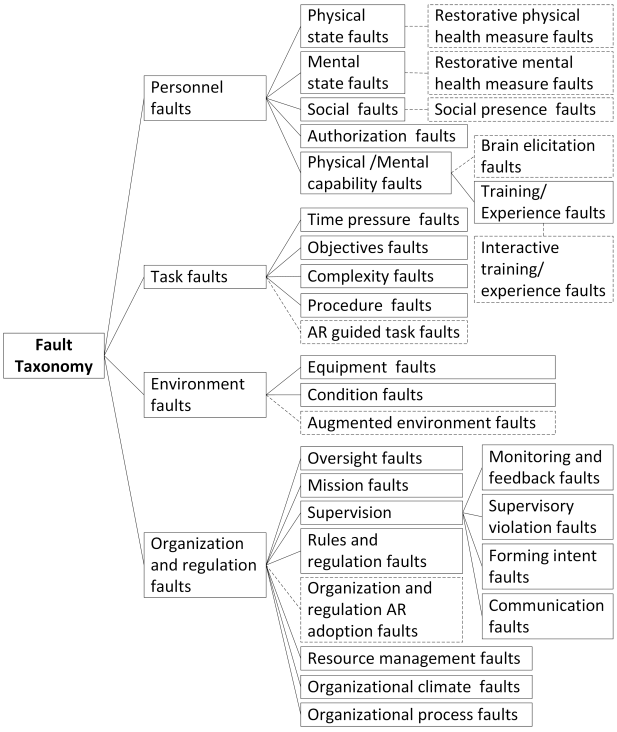

Figure 9.3: Classification of faults leading to human failures adapted from [11]

In SafeConcert metamodel, socio-technical systems can be modelled as component-based systems, where components can be software, hardware or socio entities. For socio components, which can be human or organization, the metamodel is based on SERA [8] taxonomy.

Human components are represented as composite components and subcomponents are twelve categories of human failures in SERA taxonomy. These twelve categories are divided into two types based on human functionalities (Figure 9.4). Functionalities responsible for acting (HumanActuatorUnit), with prefix "HA", including: *selection*, *response*, *knowledge decision*, *time management*, *communication*, *intent* and *feedback* and functionalities responsible for sensing (HumanSensorUnit), with prefix "HS", including: *perception*, *attention*, *sensory* and *knowledge perception*.

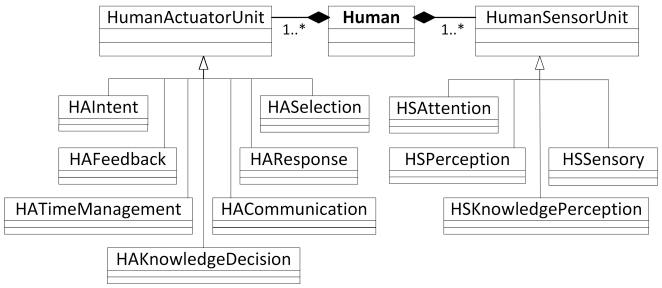

Figure 9.4: SafeConcert modeling elements to model human components [20]

Organization components are represented as composite components and subcomponents are six categories based on SERA taxonomy (Figure 9.5). These subcomponents which are called units in this metamodel are named with prefix "OU" to represent organizational unit.

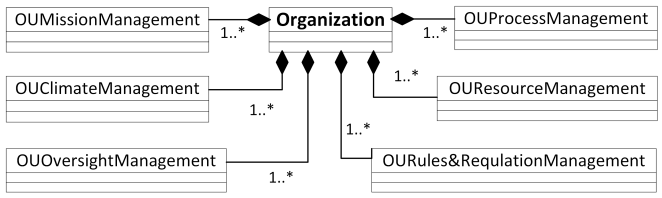

Figure 9.5: SafeConcert modeling elements to model organization components adapted from [20]

Based on SafeConcert metamodel failures are propagated from/to entities in a socio-technical system through ports and failure modes are associated to ports. Failure modes are assigned to ports by defining failure mode groups and based on domain [20].

SafeConcert is implemented in CHESS toolset [22] developed within CHESS [23] and Concerto [24] projects. This toolset offers modelling and analysis capabilities targeting high integrity systems as well as socio-technical systems. Socio entities modeling elements, which are human and organization modeling elements are based on SERA classification in this toolset. Users can define component-based architectural models composed of hardware, software, human and organization and for each component, FPTC (Failure Propagation Transformation Calculus) [25] rules (logical expressions that relate output failures to input failures) are used to model component's failure behavior. This toolset supports SafeConcert metamodel and can be extended based on the extensions provided for SafeConcert.

9.2.4 Extended SafeConcert

Based on the function classification of AR-extended humans [9], shown in Figure 9.2, we extended the human modeling elements for AR-extended human capabilities' [10]. As it is shown in Figure 9.2, there are six categories of human functions that can be divided to three types of human functionalities: functions for gaining situational awareness (SA) containing *identifying* and *paying attention*, functions for *information processing* and *deciding* and functions for *acting* and *conforming to rules*. We show these three categories by *HumanSAUnit*, *HumanProcessUnit* and *HumanActuatorUnit*. Extended human modeling elements are shown in Figure 9.6. Modeling elements that are the same as SafeConcert are shown with gray color and extended elements are shown with white color. Modeling elements characterizing AR-extended human functions are shown by dotted line border.

9.3 Extending SafeConcert

In this section, we extend SafeConcert with the aim of enabling modeling of possible faults leading to human failures including AR-stemmed faults. Some of these faults emanate from human subcomponents, which needs extension in human modeling elements and other faults emanate from organization

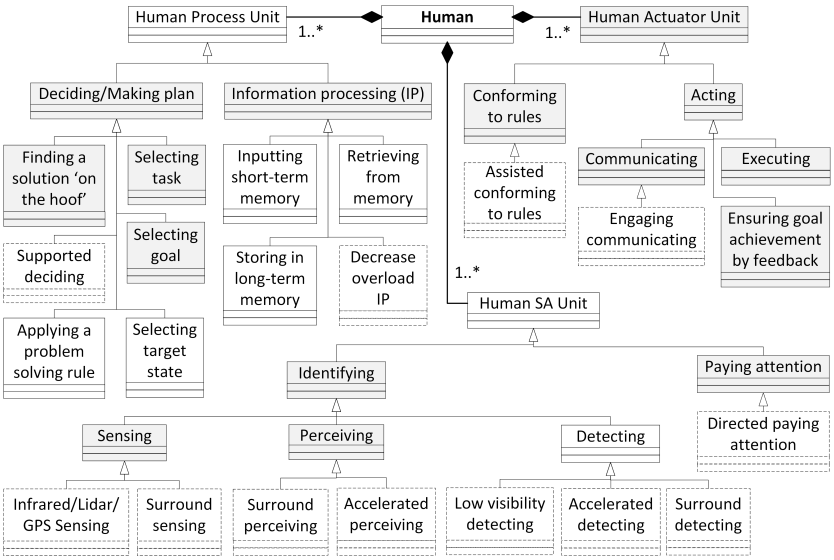

Figure 9.6: Proposed model elements to model human components [10]

subcomponents, which needs extension in organization modeling elements.

9.3.1 Extending SafeConcert Human Modeling Elements

In this subsection, we extend the human modeling elements by considering human internal faults leading to human failures. To do that, we incorporate the personnel faults in fault taxonomy shown in Figure 9.3 in human modeling elements. The result of the extension is shown in Figure 9.7. Extended modeling elements are shown with white color and AR-stemmed modeling elements are shown with dotted line border. For example, interactive training is provided by using augmented reality [26]. If there is problem in AR, this would cause failure in interactive training subcomponent, which is an internal fault for human function subcomponent and causes human failure.

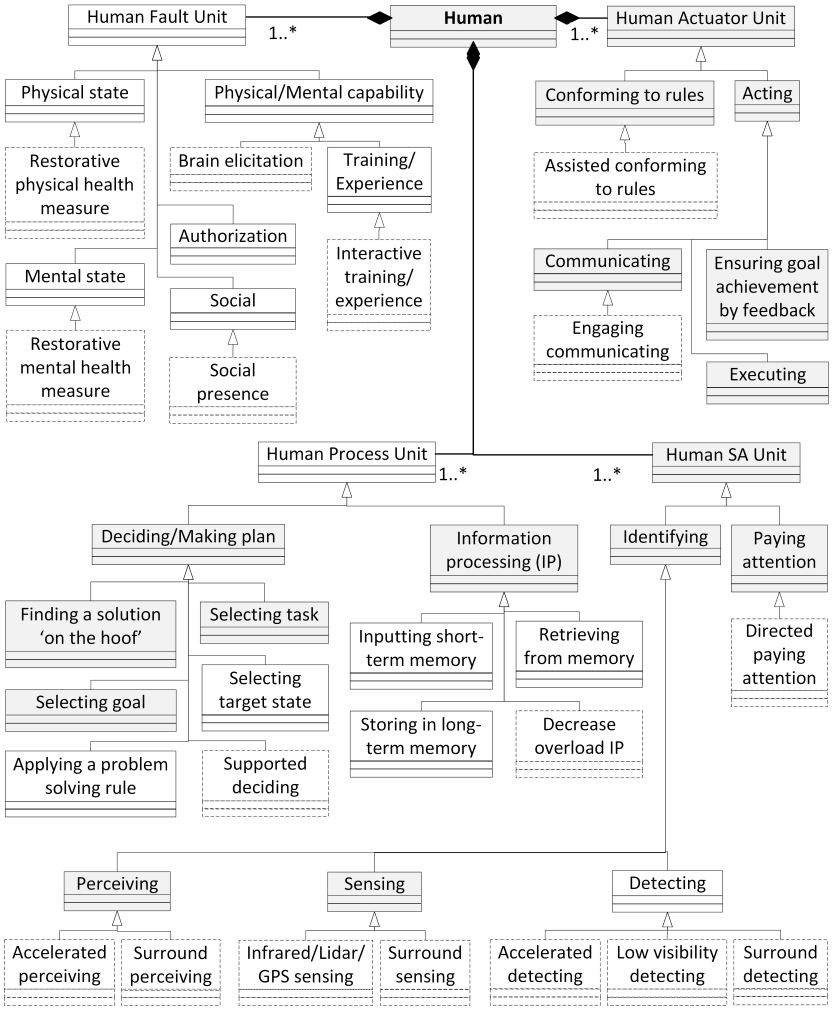

Figure 9.7: Extended model elements to model human components

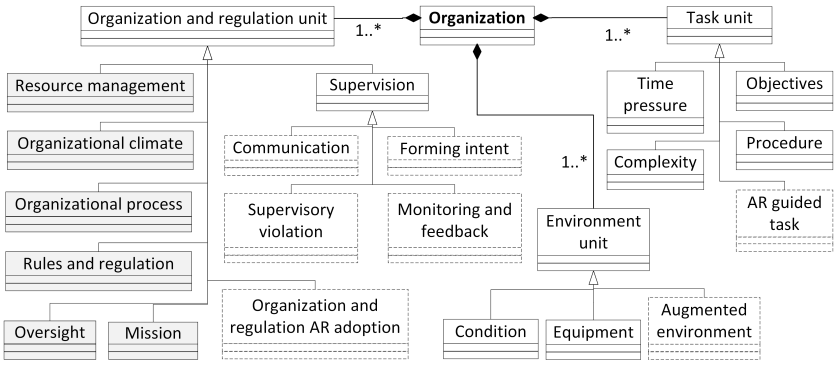

Figure 9.8: Extended model elements to model organization components

9.3.2 Extending SafeConcert Organization Modeling Elements

In this subsection, we extend organization modeling elements by considering organization, task and environment faults leading to human failures. To do that, we incorporate the organization, task and environment faults in fault taxonomy shown in Figure 9.3 in organization modeling elements. The result of the extension is shown in Figure 9.8. Extended modeling elements are shown with white color and AR-stemmed modeling elements are shown with dotted line border. For example, task procedure and environment conditions are provided by organization and their faults should be detected and corrected by organization, otherwise these faults may lead to human failures.

As it is shown in Figure 9.8, elements related to task and environment were not available in SafeConcert. Thus faults related to these categories could not be modelled either.

9.4 AR-equipped Socio-technical System Modeling

In this section, we use our extended SafeConcert to model two fictitious but credible AR-equipped socio-technical systems. The first system is AR-equipped assisted tower controlling system and the second system is AR-equipped signal passing at danger system.

9.4.1 AR-equipped Assisted Tower Controlling System Modeling

Since head down times for tower controllers could lead to catastrophic consequences, an AR tower controller assistance system is helpful for air traffic controllers (ATCOs) to provide useful information regarding air traffic and flight data projected in the front view of the aerodrome controller [4]. Development of AR displays have been taken into advisement by U.S air force to improve performance and situational awareness of ATCOs. ATCOs' duties are controlling ground traffic and air traffic within the airport traffic control area. They obtain information by observing front view through the window and using displays, patterns and other controllers. AR displays are beneficial to prevent diverting attention from front view, which is the most important source of information [27].

Within the AdCoSCo project in DLR institute, adaptive information management is combined with augmented reality to decrease information overload [4]. AR tower controller assistance system contains three main parts: context-adaptive information presentation, management of integrated information and display using augmented reality. Inputs of assistance system are from sensors and information systems. Data sources such as operator input, environment data, flight plan data and surveillance data from aerodrome surveillance ground radar are used for context-based adaptation [4].

If we consider a component-based architecture, AR tower controller assistance system, is a composite component including context-adaptive system, information management and AR display subcomponents.

A human, which is ATCO in this system, is a composite component including subcomponents from AR-extended human modeling elements. Each of the model elements in Figure 9.7, can be represented as a subcomponent. We consider physical/mental capability, deciding and acting for this system.

Civil aviation organization can be considered as organization composite component with AR adoption subcomponent. We consider AR adoption subcomponent to show the possibility of modeling AR-stemmed faults in organization.

This hypothesised model is shown in Figure 9.9. The CHESS toolset can be used to analyze the system, by defining FPTC rules. Rules are defined based on component functions and error model of them. For example, if the probability of generating failure in a subcomponent is less than a threshold and based on the related standard this failure probability is accepted, then we can assume this component will not generate fault and it may transfer the fault

from input to output, or it may detect the fault in input by fault detection techniques and prevent its propagation. Thus, defining these rules depends on each subcomponent and should be done by safety analyzer.

We assume three scenarios to show the fault propagation in model using extended modeling elements. In the first scenario (S1), there is failure in AR device, which is tower controller assistance system. Thus, the output of this component will propagate an external fault to AR-extended human, which is ATCO. This external fault would cause failure in physical/mental capability of the human and failure in deciding and finally failure in acting.

In the second scenario (S2), failure in organization component, which is civil aviation organization, will propagate an external fault to human component causing failure in physical/mental capability, deciding and acting. For example, failure in AR adoption, would cause this problem in the organization if they do not adopt AR and do not provide regulations related to AR to assist human operation. This AR adoption failure is an AR-stemmed external fault causing human failure.

In the third scenario (S3), failure in physical/mental capability of human, for example lack of required skill or attitude, is an internal fault in human component causing failure in deciding and acting.

Modeling element representing AR-caused fault is shown by gray color to illustrate the contribution of the extended modeling elements.

9.4.2 AR-equipped Signal Passing at Danger System Modeling

SPAD (signal passed at danger) is an incident when the train enters a high risk mode. There are Automatic Warning Systems (AWS) to provide an alarm for driver. We consider augmented reality alarm that provides an AR-AWS for the rail system [28].

Since driving a train is demanding, drivers have to tolerate high mental load and they should be strong in paying attention to the correct direction, for correct amount of time and with the correct priority. One of the reasons for SPADs is driver distraction or inattention. Based on a study [29] on Australian and New Zealand rail industry key factors leading to SPADs are time pressure, sighting restriction, station dwell, controller interaction and distraction.

Similar to the system considered in Subsection IV.A, in this system we need to model human, organization and technical entities. We model this socio-technical system using the extended modeling elements with gray color for AR-extended modeling elements.

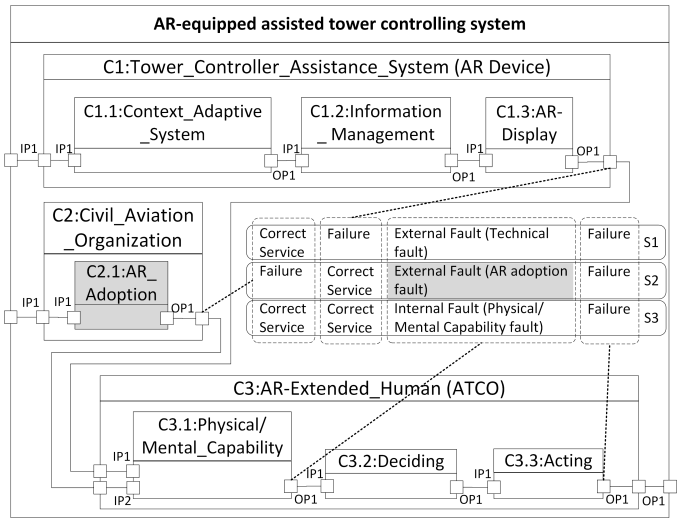

Figure 9.9: AR-equipped tower controlling system modelled with the extended SafeConcert

AR-AWS, which is a composite component within a component-based architecture representing a socio-technical system contains technical components such as AR-display.

Train driver, representing the human entity of the system is a composite component containing acting, deciding and directed paying attention functions based on the extended modeling elements in [10], and social presence based on the modeling elements extension presented in Figure 9.7.

Organization is a railway organization, which is a composite component containing objective and AR guided task subcomponents derived by elaborating on the possible behavior of the system. These two subcomponents are based on the organization extended modeling elements presented in Figure 9.8.

Similar to Subsection IV.A, we consider three scenarios depicting failure in each of the three entities causing human failures. This hypothesized model is shown in Figure 9.10.

In the first scenario (S1), there is failure in AR display, which is failure in AR-AWS and an external fault for human failure. In this scenario, AR device, which is a technical entity produces an external fault leading to human failure.

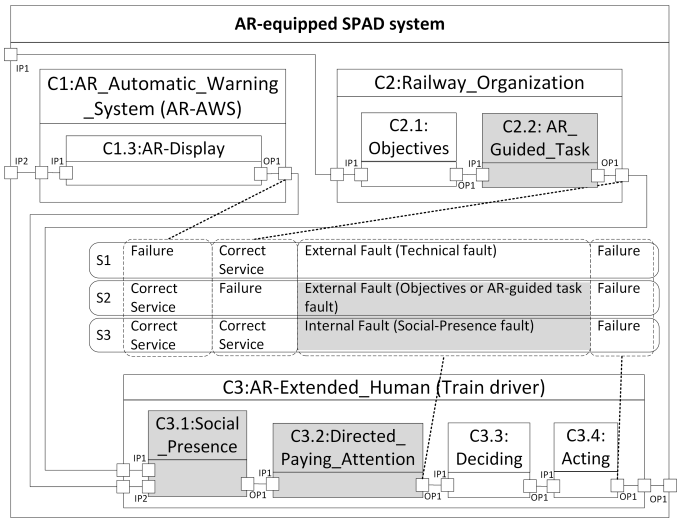

Figure 9.10: AR-equipped SPAD system modelled with the extended SafeConcert

In the second scenario (S2), AR device is working properly, but there is failure in organization. For example objectives are not defined correctly or AR guided task is not defined correctly. Thus, objective fault or AR guided task fault is an external fault leading to human failures. These two subcomponents are based on the proposed extended modeling elements in this paper (Figure 9.8).

In the third scenario (S3), AR device and organization are providing outputs without failure, while there is failure in subcomponents of the human itself, for example in social presence, which is an internal fault leading to human failure and it is based on the proposed extended modeling elements in this paper (Figure 9.7). As it is shown in this example, we can use various extended modeling elements to model internal and external faults leading to human failures, including AR-stemmed faults.

9.5 Discussion

One of the challenges in our research is that the techniques that we are extending are not used by industry and we can not use their feedback for

improvement of techniques. The traditional methods such as FTA, FMEA and FMECA are still used in most of the companies for risk assessment. These methods neither provide the explicit possibility of modeling human and AR-extended humans, nor AR-related faults leading to human failures in socio-technical systems. The issues related to AR-equipped socio-technical systems need cross-field expertise in human, AR and risk assessment and have largely remained unaddressed in techniques used in industry.

Another challenge is that augmented reality is a new technology, which is not implemented in some of the safety-critical applications that we want to have evaluation on. In addition, in some implemented cases, we do not have access to confidential information regarding the architecture of these systems to be used for evaluation. To do the risk assessment in a system, high number of scenarios with several failures of various components are required to improve safety based on this knowledge [30].

Despite the limited illustration given in Section IV, we see that our proposed extension for SafeConcert can help safety engineers during the modeling process of socio-technical systems in several ways. First, it provides the means for modeling failure behaviors of AR extended-humans and influencing factors on these failures, which are important parts of AR-equipped socio-technical systems. Second, it provides modeling elements based on human functions, AR-extended human functions and faults leading to human failures, which are in compliant with state-of-the-art human failure and faults taxonomies reviewed in [9] and [11].

There are some limitations in the proposed extension. We can not claim that human and organization components modeling elements are mutually exclusive, because sometimes it is not possible to exactly classify the human functions or organization elements involved in doing the task that are causing human failures, into one specific category and it makes the process of human and organization modeling sophisticated. Evaluation of the proposed extension is also another important issue that should be expanded to confirm its usefulness on industrial case studies.

9.6 Related works

There are several works in the literature regarding risk assessment and modeling of socio-technical systems. With the growth of utilizing new technologies in socio-technical systems, assessing the risk of using these technologies and their interaction with human in these systems is required.

In [4], authors provide risk and benefit assessment for context-adaptive augmented reality aerodrome control towers through aerodrome controllers' ratings. Several specified criteria are used for risk assessment, including transparency, complexity, interference, disruptiveness, distraction potential, failure modes and trust/complacency. Air traffic controllers were asked to rate all criteria in the range 1 to 5. Results show that context-adaptive augmented reality is helpful for controllers and improves their performance. The provided assessment is useful for demonstrating effectiveness of using augmented reality in this industry. In contrast, we try to model failure behavior of the system to overcome problems in design or implementation while developing the system.

In [31], the author proposes Safe-AR, which is a method for risk analysis of systems containing augmented reality. This method analyzes AR failures at three levels: perception, comprehension, and decision-making. To consider the safety effects of AR/user interface in risk analysis process, Safe-AR integrates failure modes related to user's mental information-processing phases. In risk assessment, likely risks and their severity are based on previous reports and the intended use of the AR. To evaluate the effectiveness of this method for other AR applications, failure modes should be generalized. In comparison to this method, our modeling method uses more general human functions and failure modes and can be considered in more AR applications.

In [32], authors propose a modelling methodology for complex socio-technical systems while new technologies are used by humans. In this method, technology modelling is used to consider its impact on system's behavior and it consists of CWA (Cognitive Work Analysis) and SD (System Dynamics) approaches [33] to capture effect of humans and dynamic interactions in complex systems. The difference of this work with ours is that the focus in this work is on complex socio-technical systems for systems engineering.

In [34], authors propose SD-BBN, which is a method that combines Bayesian belief networks (BBN) [35] and system dynamics (SD) [36] for socio-technical predictive modeling. In BBN, probabilities of causes and effects are shown by conditional probabilities. Expert opinion is used for defining the probabilities. To consider feedback loops and dynamic interactions of causal factors, this method combines BBN with SD. SD is a simulation-based modeling technique that is useful for modeling organizational behavior, dynamics and feedback. This SD-BBN method is integrated with classical probabilistic risk analysis (PRA) [37] techniques and fault tree and event tree are used to model system risk. This model is used to

predict happening of accidents in a period of time and guide managers to schedule their activities, while our model is used during the system development process for eliminating design failures incrementally and iteratively.

9.7 Conclusion

In this paper, we performed an additional step towards assessing risk of safety-critical socio-technical systems containing augmented reality. As known, risk assessment starts with the identification of what can go wrong. Our previously proposed human failure and fault taxonomies may act as helpful means during the identification by offering AR-specific keywords. Their coherent incorporation (proposed in this paper) within SafeConcert, a metamodel targeting socio-technical systems, helps in getting the component-level view of what can go wrong and enables compositional analysis tools to calculate what can go wrong at system level. We illustrated our extension on two fictitious but still credible systems from air traffic control and rail domains.

As future work, we aim at implementing the conceptual extension of SafeConcert within CHESSML [21]. In addition, we aim at extending current compositional analysis techniques to be able to calculate what can go wrong at system level. Specifically, our starting point will be Concerto-FLA [38], which is a plugin within the CHESS toolset, part of the, recently released, open-source AMASS platform for certification [39].

Acknowledgment

This work is funded by EU H2020 MSC-ITN grant agreement No 764951.

Bibliography

[1] Van Krevelen, D., Poelman, R.: A survey of augmented reality technologies, applications and limitations. The International Journal of Virtual Reality **9**(2) (2010) 1–20

[2] Goldiez, B.F., Saptoka, N., Aedunuthula, P.: Human performance assessments when using augmented reality for navigation. Technical report, University of Central Florida Orlando Inst for Simulation and Training (2006)

[3] Goldiez, B.F., Ahmad, A.M., Hancock, P.A.: Effects of augmented reality display settings on human wayfinding performance. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) **37**(5) (2007) 839–845

[4] Gürlük, H., Gluchshenko, O., Finke, M., Christoffels, L., Tyburzy, L.: Assessment of risks and benefits of context-adaptive augmented reality for aerodrome control towers. In: Digital Avionics Systems Conference (DASC), IEEE (2018) 1–10

[5] Avizienis, A., Laprie, J.C., Randell, B., Landwehr, C.: Basic concepts and taxonomy of dependable and secure computing. IEEE transactions on dependable and secure computing **1**(1) (2004) 11–33

[6] Yeh, M., Wickens, C.: Attention and trust biases in the design of augmented reality displays. University of Illinois at Urbana-Champaign, Aviation Research Lab (2000)

[7] Dowell, S.R., Foyle, D.C., Hooey, B.L., Williams, J.L.: The effect of visual location on cognitive tunneling with superimposed hud symbology. In: Proceedings of the human factors and ergonomics society annual

meeting. Volume 46., SAGE Publications Sage CA: Los Angeles, CA (2002) 121–125

[8] Hendy, K.C.: A tool for human factors accident investigation, classification and risk management. Technical report, Defence Research And Development Toronto (Canada) (2003)

[9] Sheikh Bahaei, S., Gallina, B.: Augmented reality-extended humans: towards a taxonomy of failures – focus on visual technologies. In: European Safety and Reliability Conference (ESREL), Research Publishing, Singapore (2019)

[10] Sheikh Bahaei, S., Gallina, B.: Towards Assessing Risk of Reality Augmented Safety-critical Socio-technical Systems. Published as proceedings annex on the International Symposium on Model-Based Safety and Assessment (IMBSA) website. http://easyconferences.eu/imbsa2019/proceedings-annex/ (2019)

[11] Sheikh Bahaei, S., Gallina, B., Laumann, K., Rasmussen Skogstad, M.: Effect of augmented reality on faults leading to human failures in socio-technical systems. In: International Conference on System Reliability and Safety (ICSRS), IEEE (2019)

[12] Norman, D.A.: Errors in human performance. Technical report, California Univ San Diego LA JOLLA Center For Human Information Processing (1980)

[13] Reason, J.: The human contribution: unsafe acts, accidents and heroic recoveries. CRC Press (2017)

[14] Rasmussen, J.: Human errors. a taxonomy for describing human malfunction in industrial installations. Journal of occupational accidents **4**(2-4) (1982) 311–333

[15] Shappell, S.A., Wiegmann, D.A.: The human factors analysis and classification system–HFACS. Technical report, Civil Aeromedical Institute (2000)

[16] Stanton, N.A., Salmon, P.M.: Human error taxonomies applied to driving: A generic driver error taxonomy and its implications for intelligent transport systems. Safety Science **47**(2) (2009) 227–237

[17] Phan, M.T.: Estimation of driver awareness of pedestrian for an augmented reality advanced driving assistance system. PhD thesis, Université de Technologie de Compiègne (2016)

[18] Gertman, D., Blackman, H., Marble, J., Byers, J., Smith, C., et al.: The SPAR-H human reliability analysis method. US Nuclear Regulatory Commission **230** (2005)

[19] Miller, M.R., Jun, H., Herrera, F., Villa, J.Y., Welch, G., Bailenson, J.N.: Social interaction in augmented reality. PloS one **14**(5) (2019) e0216290

[20] Montecchi, L., Gallina, B.: SafeConcert: A metamodel for a concerted safety modeling of socio-technical systems. In: International Symposium on Model-Based Safety and Assessment, Springer (2017) 129–144

[21] CONCERTO D2.7 – Analysis and back-propagation of properties for multicore systems – Final Version: http://www.concerto-project.org/results

[22] Cicchetti, A., Ciccozzi, F., Mazzini, S., Puri, S., Panunzio, M., Zovi, A., Vardanega, T.: Chess: a model-driven engineering tool environment for aiding the development of complex industrial systems. In: Proceedings of the 27th IEEE/ACM International Conference on Automated Software Engineering, ACM (2012) 362–365

[23] ARTEMIS-JU-100022 CHESS – Composition with guarantees for high-integrity embedded software components assembly: http://www.chess-project.org/

[24] ARTEMIS-JU-333053 CONCERTO – Guaranteed Component Assembly with Round Trip Analysis for Energy Efficient High-integrity Multi-core systems: http://www.concerto-project.org

[25] Wallace, M.: Modular architectural representation and analysis of fault propagation and transformation. Electronic Notes in Theoretical Computer Science **141**(3) (2005) 53–71

[26] Lee, K.: Augmented reality in education and training. TechTrends **56**(2) (2012) 13–21

[27] Ruffner, J.W., Fulbrook, J.E.: Usability considerations for a tower controller near-eye augmented reality display. In: Proceedings of the

Human Factors and Ergonomics Society Annual Meeting. Volume 51., Sage Publications Sage CA: Los Angeles, CA (2007) 117–121

[28] Naweed, A., Rainbird, S., Chapman, J.: Investigating the formal countermeasures and informal strategies used to mitigate spad risk in train driving. Ergonomics **58**(6) (2015) 883–896

[29] Naweed, A., Rainbird, S.: Risk factors moderating driving-related distraction and inattention in the natural rail environment. In: 3rd International Conference on Driver Distraction and Inattention. (2013)

[30] Zio, E.: The future of risk assessment. Reliability Engineering and System Safety **177** (September 2018) 176–190

[31] Lutz, R.R.: Safe-AR: Reducing risk while augmenting reality. In: 2018 IEEE 29th International Symposium on Software Reliability Engineering (ISSRE), IEEE (2018) 70–75

[32] Oosthuizen, R., Pretorius, L.: Assessing the impact of new technology on complex sociotechnical systems. South African Journal of Industrial Engineering **27**(2) (2016) 15–29

[33] Oosthuizen, R., Pretorius, L.: Modelling methodology for engineering of complex sociotechnical systems. In: INCOSE International Symposium. Volume 24., Wiley Online Library (2014) 268–281

[34] Mohaghegh, Z.: Combining system dynamics and bayesian belief networks for socio-technical risk analysis. In: 2010 IEEE International Conference on Intelligence and Security Informatics, IEEE (2010) 196–201

[35] Pearl, J.: Bayesian networks: A model of self-activated memory for evidential reasoning. In: Proceedings of the 7th Conference of the Cognitive Science Society, 1985. (1985) 329–334

[36] Sterman, J.: Business Dynamics: System Thinking and Modeling for a Complex World. Irwin McGraw-Hill (2000)

[37] Bedford, T., Cooke, R., et al.: Probabilistic risk analysis: foundations and methods. Cambridge University Press (2001)

[38] Gallina, B., Sefer, E., Refsdal, A.: Towards safety risk assessment of socio-technical systems via failure logic analysis. In: 2014 IEEE International Symposium on Software Reliability Engineering Workshops, IEEE (2014) 287–292

[39] AMASS: Architecture-driven, Multi-concern and Seamless Assurance and Certification of Cyber-Physical Systems. http://www.amass-ecsel.eu/

Chapter 10

Paper D: A Case Study for Risk Assessment in AR-equipped Socio-technical Systems

Soheila Sheikh Bahaei, Barbara Gallina and Marko Vidović
Technical Report, ISRN MDH-MRTC-332/2020-1-SE, Mälardalen Real-Time Research Center, Mälardalen University, May 2020.

Abstract

Augmented Reality (AR) technologies are used as human-machine interface within various types of safety-critical systems. In order to avoid unreasonable risk, it is required to anticipate new types of dependability threats (faults, errors, failures), which could be introduced within the systems by these technologies. In our previous work, we have designed an extension for CHESS framework to capture AR-related dependability threats (focusing on faults and failures) and we have extended its metamodel, which provides qualitative modeling and analysis capabilities that can be used for AR-equipped socio-technical systems. In this paper, we conduct a case study from automotive domain to present modeling and analysis capabilities of our proposed extensions. We conduct qualitative modeling and analysis based on Concerto-FLA analysis technique, which is an analysis technique for socio-technical systems to find out if the proposed extensions would be helpful in capturing new system failures caused by AR-related dependability threats.

10.1 Introduction

Augmented Reality (AR) technology is used for superimposing virtual and computer generated information on the reality of the user [1]. This information would be visual, auditory, etc., for enhancing human capabilities [2]. An example of visual augmented reality is using navigational information superimposed on the windshield of a car for driver guidance.

Utilizing augmented reality technology in socio-technical systems demands risk assessment to make sure that it is not harmful for people and the environment, while interacting with humans. Thus, it is required to identify the threats and their propagation via modeling the system and analyzing its behavior in order to enable risk analysis of systems containing augmented reality.

According to ISO 26262 [3] standard, the automotive standard for functional safety, risk assessment is a "method to identify and categorize hazardous events of items and to specify safety goals and ASILs (Automotive Safety Integrity Level) related to the prevention or mitigation of the associated hazards in order to avoid unreasonable risk". In order to identify AR-related hazardous events or dependability threats, which are risk sources, we have proposed two taxonomies in our previous works. Based on these taxonomies extensions are provided to investigate AR-related dependability threats in architecture modeling and analysis. So far, however, there has been little investigation about how effective are current modeling and analysis techniques for industrial systems containing new technologies and if it is possible to capture risk caused by augmented reality-related new threats.

In this paper, we use an industrial case study for evaluating our proposed conceptual extensions on CHESS framework for capturing AR-related dependability threats in AR-equipped socio-technical systems. Conceptual extensions are mostly associated with SafeConcert metamodel [4], which is part of the modeling language included in the CHESS framework for modelling socio-technical systems. Extended metamodel provides modeling and analysis capabilities. In order to show the analysis capabilities of the proposed extensions, we use Concerto-FLA [5], which is an analysis technique for socio-technical systems. Concerto-FLA uses Fault Propagation and Transformation Calculus (FPTC) [6] syntax to provide the means for analysis in system level. We present the case study based on SEooC (Safety element out of context) concept of ISO 26262 standard, which refers to elements that are not developed in the context of a particular vehicle. Based on this concept, assumptions should be defined for the context in which a

component is going to be used [7]. Finally, we provide threats to validity and limitations and benefits of the extensions.

The rest of the paper is organized as follows. In Section 10.2, we provide essential background information. In Section 10.3, we design and conduct the case study to evaluate modeling and analysis capabilities of the proposed extensions and we discuss about lessons learnt based on limitations and benefits of our research. In Section 10.4, we discuss about threats to validity in relation to our research. Finally, in Section 10.5, we present some concluding remarks and sketch future work.

10.2 Background

This section provides essential background information onto which our work is based. First, CHESS framework is introduced. Then, SafeConcert modelling technique and AR-related modeling extensions are presented. FPTC syntax and Concerto-FLA analysis technique are also explained. Finally, ISO 26262, SEooC and SAE automation levels are presented.

10.2.1 CHESS Framework

CHESS framework [8] provides a methodology, a language and a toolset for developing high-integrity systems.

The CHESS methodology, which is component-based and model-driven, is based on an incremental and iterative process. Based on this methodology, components are defined incrementally with functional and also extra-functional properties, such as dependability information [9]. Then, developers can use the analysis and back propagate the results iteratively.

CHESS-ML (CHESS Modeling Language) [10] is based on UML and provides the modeling elements required for modeling high-integrity systems.

CHESS toolset includes a set of plugins for code generation and providing various analysis capabilities. For example, Failure Logic Analysis (FLA) is a plugin related to analysis. In FLA, component-based model of the system is provided and dependability information is used for decorating components. Then, analysis results can be back propagated to the system model.

In this paper, our focus is on failure logic analysis and we consider Concerto-FLA as the analysis technique used in this toolset. Concerto-FLA is based on Fault Propagation and Transformation Calculus (FPTC) [6].

10.2.2 SafeConcert and Its Extension of AR

SafeConcert [4] is a metamodel for modeling socio and technical entities in socio-technical systems. This metamodel is part of CHESS-ML modeling language [10], which is a UML-based modeling language. In SafeConcert metamodel, software, hardware or socio entities can be modelled as components in component-based systems representing socio-technical systems. SERA taxonomy [11] is used for modeling human and organization, which are the socio entities of the system. In this metamodel human sub-components are modelled based on twelve categories of human failures including failures in perception, decision, response, etc.

In [12], we extended human modeling elements based on AREXTax, which is an AR-extended human function taxonomy [13] gained by harmonizing about 6 state-of-the-art human failure taxonomies (Norman [14], Reason [15], Rasmussen [16], HFACS (Human Factor Analysis and Classification System) [17], SERA (Systematic Error and Risk Analysis) [11], Driving [18]) and then extending the taxonomy based on various studies and experiments on augmented reality. These extended modeling elements are shown in Figure 10.1. Human functions are divided to three categories including human process unit, human SA unit, and human actuator unit. Human fault unit are related to human internal influencing factors on human function. This part will be explained in next paragraph. Extended modeling elements are shown with white color and AR-stemmed modeling elements are shown with dotted line border. These extended modeling elements enable modeling of AR-extended human functions. For example, detection failure, which is failure in detecting human function, is a human failure introduced by several human failure taxonomies such as Reason [15] and Rasmussen[16] taxonomies. Based on experiments and studies on augmented reality including [19] and [20], detecting function would be extended to surround detecting while using AR (surrounding information would be augmented on real world view of the user by AR), thus surround detecting can be considered as an extended sub-component of human component, which is an extended modeling element proposed for analysis of AR-equipped socio-technical systems.

In [21], we extended organization modeling elements based on AREFTax, which is a fault taxonomy including AR-caused faults [22] gained by harmonizing about 5 state-of-the-art fault taxonomies (Rasmussen [16], HFACS (Human Factor Analysis and Classification System) [17], SERA (Systematic Error and Risk Analysis) [11], Driving [18] and SPAR-H

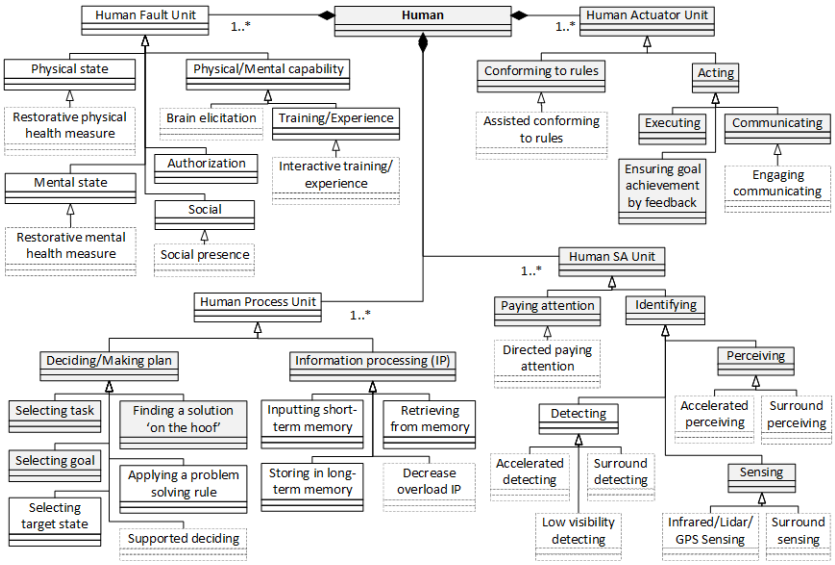

Figure 10.1: Extended modeling elements for human components [21]

(Standardized Plant Analysis Risk Human Reliability Analysis)[23]) and then extending the taxonomy based on various studies and experiments on augmented reality. These extended modeling elements are shown in Figure 10.2 and human fault unit of Figure 10.1. Extended modeling elements are shown with white color and AR-stemmed modeling elements are shown with dotted line border. These extended modeling elements enable modeling of AR-caused faults leading to human failures. Faults would be caused by human, environment, organization, etc. Human related faults are categorized in human fault unit of Figure 10.1 and non-human faults are categorized as three categories of organizational factors including organization and regulation unit, environment unit and task unit. For example, failure in physical state of a human is a human internal fault leading to human failure. This is shown as human modeling element in human fault unit category shown in Figure 10.1. Another example is condition, which is a non-human fault leading to human failure and it is categorized in organization taxonomy shown in Figure 10.2. One example of the AR-extended modeling elements is social presence shown in Figure 10.1. Based on studies on augmented reality [24], using AR would decrease social presence and failure in social presence

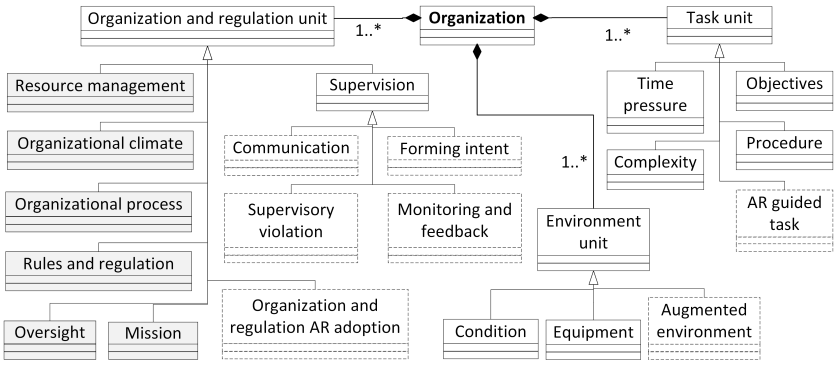

Figure 10.2: Extended modeling elements for organization components [21]

can be considered as fault leading to human failure.

10.2.3 The FPTC Syntax

FPTC syntax was proposed as part of FPTC analysis technique [6]. FPTC rules are set of logical expressions that relate output failure modes to combinations of input failure modes in each individual component [25].

Components' behavior can be classified as source (if component generates a failure), sink (if component is able to detect and correct input failure), propagational (if component propagates failures received in its input to its output) and transformational (if component transforms the type of failure received in its input to another type in its output).

FPTC syntax for modeling failure behavior at component and connector level is as follows:

behavior = expression+
expression = LHS '→' RHS
LHS = portname'.' bL | portname '.' bL (',' portname '.' bL) +
RHS = portname'.' bR | portname '.' bR (',' portname '.' bR) +
failure = 'early' | 'late' | 'commission' | 'omission' | 'valueSubtle' | 'valueCoarse'
bL = 'wildcard' | bR
bR = 'noFailure' | failure

Early and late failures refer to provided function at a wrong time (early or late). Commission failures refer to provided function at a time which is not

expected and omission failures refer to not provided function at a time which is expected. Value failures refer to wrong value after computations, which would be valueSubtle (user can not detect it) or valueCoarse (user can detect it).

Wildcard in an input port shows that the output behavior is the same regardless of the failure mode on this input port. NoFailure in an input port shows normal behavior.

Based on this syntax, "IP1.noFailure → OP1.omission" shows a source behavior and should be read as follows: if the component receives noFailure (normal behavior) on its input port IP1, it generates omission on its output port OP1.

10.2.4 Concerto-FLA Analysis Technique

Concerto-FLA [5], which extends FPTC [6], is a model-based analysis technique that provides the possibility for analyzing failure behavior of humans and organizations in addition to technical entities by using SERA [11] classification of socio-failures. As we recalled in Subsection 10.2.1, this technique is provided as a plugin within the CHESS toolset and allows users to define component-based architectural models composed of hardware, software, human and organization. This technique includes five main steps.

1. Modeling architectural elements including software, hardware, human, organization, connectors, interfaces and etc.

2. Modeling failure behavior at component and connector level by using FPTC syntactical rules. Concerto-FLA has adopted the FPTC syntax for modeling failure behavior at component and connector level (explained in Subsection 10.2.3).

3. Modeling failure modes at system level by injection of inputs.

4. Performing qualitative analysis through automatic calculation of the failure propagations. This step is similar to FPTC technique that system architecture is considered as a token-passing network and set of possible failures that would be propagated along a connection is called tokenset (default value for each tokenset is noFailure, which means normal behavior). In order to obtain system behavior, maximal tokenset is calculated for each connection through a fixed-point calculation.

5. Interpreting the results at system level. Based on the interpretation it will be decided to do the re-design or not.

10.2.5 ISO 26262, SEooC and SAE Automation Levels

ISO 26262 standard [3] provides the requirements and set of activities that should be performed during the lifecycle phases such as development, production, operation, service and decommissioning. Integrity level or ASIL (Automotive Safety Integrity Levels) are determined and used for applying the requirements to avoid unreasonable residual risk. ASIL specifies item's necessary safety requirements to achieve an acceptable residual risk. Residual risks are remaining risks after using safety measures.

Safety element out of context (SEooC), introduced by ISO 26262, refers to an element that is not defined in the context of a special vehicle, but it can be used to make an item, which implements functions at vehicle level. SEooC is based on ISO 26262 safety process and information regarding system context such as interactions and dependencies on the elements in the environment should be assumed [26].

The SEooC development contains 4 main steps:

1. (a) Definition of the SEooC scope: assumptions related to the scope, functionalities and external interfaces of the SEooC should be defined.
   (b) Definition of the assumptions on safety requirements for the SEooC: assumptions related to item definition, safety goals of the item and functional safety requirements related to SEooC functionality required for defining technical safety requirements of the SEooC should be defined.
2. Development of SEooC: based on the assumed functional safety requirements, technical safety requirements are derived and then SEooC is developed based on ISO 26262 standard.
3. Providing work products: work products are documents that show the fulfilled functional safety requirements and requirements and assumptions on the context of SEooC.
4. Integration of the SEooC into the item: safety goals and functional safety requirements defined in item development should match with assumed functional safety requirements for the SEooC. In case of a SEooC assumption mismatch, change management activity based on ISO 26262 standard should be conducted.

Safety process of the ISO 26262 standard, shown in Figure 10.3, starts with *concept phase* containing *item definition*, *hazard analysis and risk*

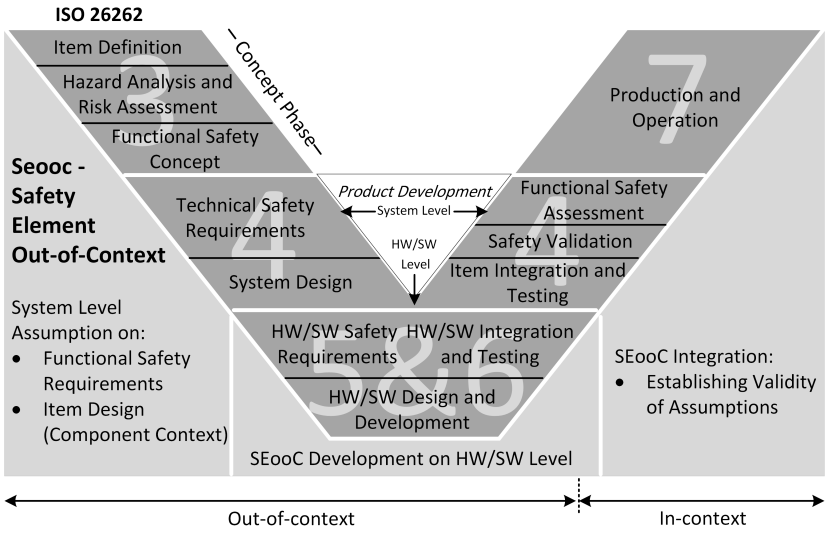

Figure 10.3: Alignment of the ISO 26262 lifecycle activities to SEooC development and integration [26]

assessment and *functional safety concept* [26]. An *item* implements a vehicle level function. In *item* definition the main objective is defining items, which requires defining the dependencies and interactions with environment. Then, related hazards are identified and functional safety requirements are obtained. In SEooC, assumptions related to system context are the main output of the *concept phase* sent to the *product development phase*. *Product development phase* contains *system level* and *HW/SW level*. Functional safety concept is used to provide technical safety requirements and to design system in *product development phase* at system level. Then, hardware and software development and testing are done based on system design. HW/SW safety requirements are based on assumptions provided in concept phase. Next step in the process is integration and testing of HW/SW elements and then in system level integration of elements that compose an item, safety validation and functional safety assessment are done, which require establishing validity of assumptions. Finally, the last step is production and operation.

Based on the taxonomy and definitions related to driving automation systems for on-road motor vehicles performing part or the entire dynamic

driving task (DDT) on a sustained basis, there are six levels of driving automation. SAE level 0 refers to no driving automation and SAE level 5 refers to full driving automation [27]. Assessing human factor in driver-vehicle interface is not only important on lower SAE levels, but also on higher levels because of the importance of safe transition between automated and non-automated vehicle operation [28]. In order to improve safety, various scenarios of driver/vehicle interaction should be considered.

10.3 Case Study Design and Execution

In this section, we design a case study to present the modeling and analyzing capabilities of proposed extensions for CHESS framework that can be used to qualitatively analyze the emerging risks for AR-equipped socio-technical systems.

Alignment of the risk assessment activities to the ISO 26262 development process is shown in Figure 10.4. There are four main steps. The first step is to define composite components of the system. In order to find composite components, we need to answer to the question of what are the involved entities. Second step is to determine sub-components of each composite component. In order to determine sub-components, we need to identify different effective aspects of each entity. In this step, our proposed taxonomies and extended modeling elements can be helpful to provide a list of effective aspects and based on scenario and the selected case study, required sub-components can be selected. Third step is to model the behavior of each sub-component, which should be done based on analysis of each sub-component individually. In order to model each sub-component behavior, effect of related aspect to the sub-component's behavior should be identified. Finally, last step is analyzing system behavior, which provides effect of various aspects on the system.

10.3.1 Objectives

Our objectives include presenting the modeling capabilities and analysis capabilities of our proposed AR-related extensions in order to estimate how effective they are in predicting new kinds of risks caused by AR-related factors. In order to do that, we use an industrial case study from automotive domain to evaluate the proposed extensions.

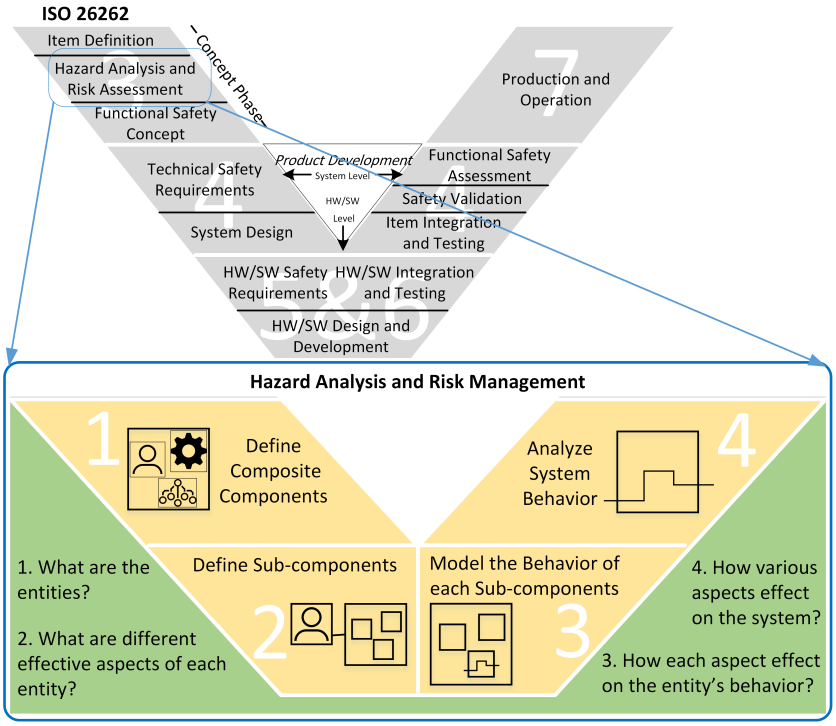

Figure 10.4: Alignment of the risk assessment activities to ISO 26262 development process

10.3.2 Research Methodology

We use case study research methodology based on [29]. The steps carried out for the presented research is presented in Figure 10.5. In the first step, the first and second authors discussed about objectives and the structure of the research.

In the second step, the first and second authors asked from Xylon Company for a case study in the context of augmented reality socio-technical systems and third author suggested surround view system as a case study and a meeting was organized between three authors to decide about the collaboration. First and third authors also discussed about system description.

In the third step, system architecture was provided by the first author based on information provided by third author and it was reviewed by third author in

some iterations for improvement. Second author also reviewed the architecture and provided comments and suggestions for improvement.

In the fourth step, analysis of the case study was provided by the first author based on Concerto-FLA analysis technique and it was reviewed by the second and third authors.

In the fifth step, the first author provided discussion about results and second and third authors reviewed the results. Second author also provided suggestions for improvement and for discussing about validity of the work.

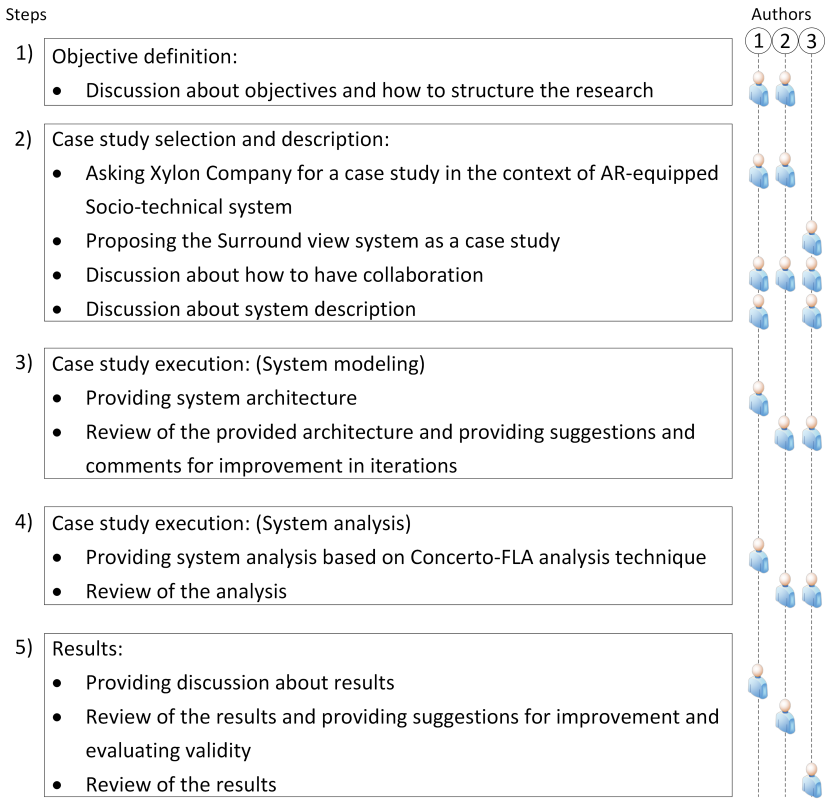

Figure 10.5: Steps taken for the carried out research

10.3.3 Case Study Selection and Description

The case study is conducted in collaboration with Xylon, an electronic company providing intellectual property in the fields of embedded graphics, video, image processing and networking.

In this study, we select as case study subject a socio-technical system containing the following entities:

- Road transport organization (socio entity): representing the organization responsible for providing transport rules and regulations, proper road conditions and etc.
- Driver (socio entity): representing a human who is expected to drive a vehicle and park it safely by utilizing augmented reality technology used in the surround view system of the vehicle.
- Vehicle (technical entity): representing vehicle containing surround view system (a SEooC of driving automation level 0 that includes augmented reality technology used to empower drivers).

Surround view systems are used to assist drivers to park more safely by providing a 3D video from the surrounding environment of the car. In Figure 10.6, it is illustrated how the 3D video is shown to the driver. As it is shown in Figure 10.6, driver can have a top view of the car while driving. This top view is obtained by compounding 4 views captured by 4 cameras mounted around the car and by changing point of view. It is like there is a flying camera visualizing vehicle's surrounding, which is called virtual flying camera feature. A picture of a virtual car is also augmented to the video to show the position of the car. Navigation information and parking lines also can be annotated to the video by visual AR technology. The current surround view system is a SEooC of driving automation level 0. However, Xylon plan to develop automated driving system features to higher levels for the future versions of the system.

Assumptions on the scope of the SEooC are:

- The system can be connected to the rest of the vehicle in order to obtain speed information. In case of drawing parking path lines, steering wheel angle and information from gearbox would also be obtained to determine reverse driving.

Assumptions on functional requirements of the SEooC are:

Figure 10.6: Sample images from 3D videos provided in surround view system

- The system is enabled either at low speed or it can be activated manually by the driver.
- The system is disabled either when moving above some speed threshold or it can be deactivated by driver.

Assumptions on the functional safety requirements allocated to the SEooC are:

- The system does not activate the function at high vehicle speed automatically.
- The system does not deactivate the functionality at low speed automatically.

10.3.4 Case Study Execution: System Modelling

This subsection reports how we model the described system in Subsection 10.3.3 using our proposed extensions. Subsection 10.3.3 provides the required information for the first step of the risk assessment process defined in Figure 10.4, which is identifying the entities for defining composite components. Based on the selected case study explained in Subsection 10.3.3, organization, driver and a vehicle containing an automotive surround view system are three composite components of this system. In this subsection we provide information for the second and third steps of risk assessment process. Effective aspects of each entity, which are modeled as sub-components of

each composite component are selected from AREXTax and AREFTax explained in Subsection 10.2.2, for socio entities and based on system description for vehicle, which is a technical entity.

- Road transport organization effective aspects:
  - Organization and regulation AR adoption: it refers to upgrading rules and regulations of road transport organization based on AR technology.
  - Condition: it refers to road condition.
  - AR guided task: it refers to the task, which AR is used for guiding driver to do that. For example, if AR is used to guide driver to park the car more safely, parking safely is the AR-guided task.
- Driver effective aspects:
  - Surround detecting: it refers to an AR-extended function, because driver can detect surround environment through AR technology.
  - Supported deciding: it refers to an AR-extended function, because driver can decide with the support of AR technology.
  - Executing: it refers to human executing function.
  - Interactive experience: it refers to an AR-caused factor, because AR provides interactive ways for enhancing user experience.
  - Social presence: it refers to an AR-caused factor, because AR may decrease social presence and lead to human failure.
- Vehicle containing surround view system effective aspects:
  - A set of speed sensors: each sensor is a hardware for providing speed of the vehicle based on its movement.
  - A set of cameras: each camera is a hardware for providing raw data for a video receiver. Usually there are four cameras that can be attached to four sides of the car.
  - Switch: switch is a hardware for receiving on/off command from driver. It is also possible to send on/off command automatically based on driving requirement.
  - Peripheral controller: peripheral controller includes hardware and software for receiving user inputs such as on/off command and speed and for sending them to user application implementation.

- A set of video receivers: each video receiver includes a hardware and a driver. Its hardware is used for transforming raw data to AXI-stream based on the command from its driver implementation.
- Video storing unit: video storing unit includes a hardware and a driver. Its hardware is used for receiving AXI-stream and storing it to the memory by means of DDR memory controller based on the command received form its driver.
- DDR controller: DDR controller is a hardware for accessing DDR memory, which stores video in DDR memory and provides general memory access to all system IPs.
- Video processing IP: Video processing IP includes hardware and software for reading prepared data structures and video from memory and for processing video accordingly and finally for storing the processed video to memory through DDR controller. The prepared data is stored to memory by video processing IP driver based on the data structures received from memory.
- Display controller: Display controller includes hardware and software for reading memory via DDR memory controller where processed video is stored and for converting it in the format appropriate for driving displays.
- Processing unit: processing unit includes hardware and software, which its software contains all the software and drivers of all other IPs. The software also contains user application implementation and video processing engine implementation. User application implementation receives inputs from peripheral unit and controls operation of all IPs by means of their software drivers. Video processing engine implementation prepares data structures to be stored in DDR memory through DDR controller.

Figure 10.7 provides an overview of the integration of the human, organization and vehicle effective aspects.

In Figure 10.8, we show how this AR-equipped socio-technical system is modeled. Driver is composed of five sub-components. Driver has four inputs and one of its inputs is from system input with the name human detection input (HDI). Two other inputs are from organization and surround view system and the last input is human communication input (HCI). We consider also interactive experience and social presence as two sub-components of human component, which are influencing factors on human functions.

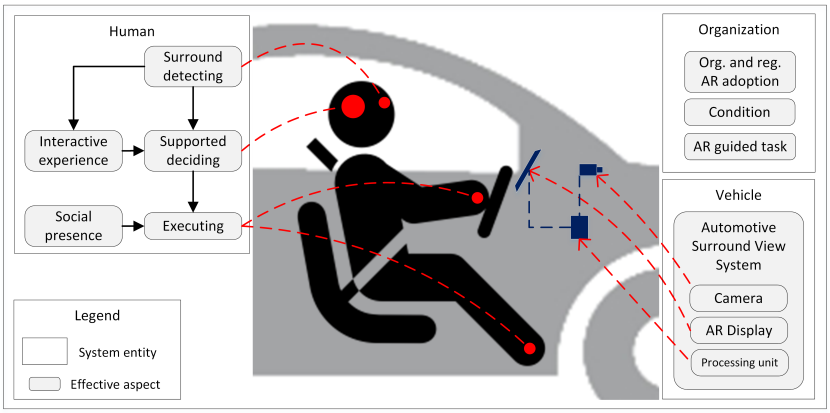

Figure 10.7: Integration of the human functions and influencing factors with SEooC

Interactive experience effects on supported deciding and is effected by surround detecting. Social presence receives input from system with the name human communication input (HCI) and effects on human executing. Driver output, which is output of the system is human action shown by HA.

Organization and regulation AR adoption, condition and AR guided task are three sub-components of organization composite component. Organization component receives input from system, which represents influences from regulation authorities on the organization (REG). Human, organization and their relation with surround view system are modelled in Figure 10.8. Gray color is used to show the extended modelling elements used in this system.

Vehicle is also modelled with three inputs including user command shown by CMD, vehicle movement shown by VMV and camera input shown by CAM.

10.3.5 Case Study Execution: System Analysis

This subsection reports on the analysis of the system using our proposed extensions, which refers to the last step of the risk assessment process defined in Figure 10.4. We follow the five steps of Concerto-FLA analysis technique explained in Subsection 10.2.4 for system analysis.

1. First step is provided in Figure 10.8. We explained how the system is

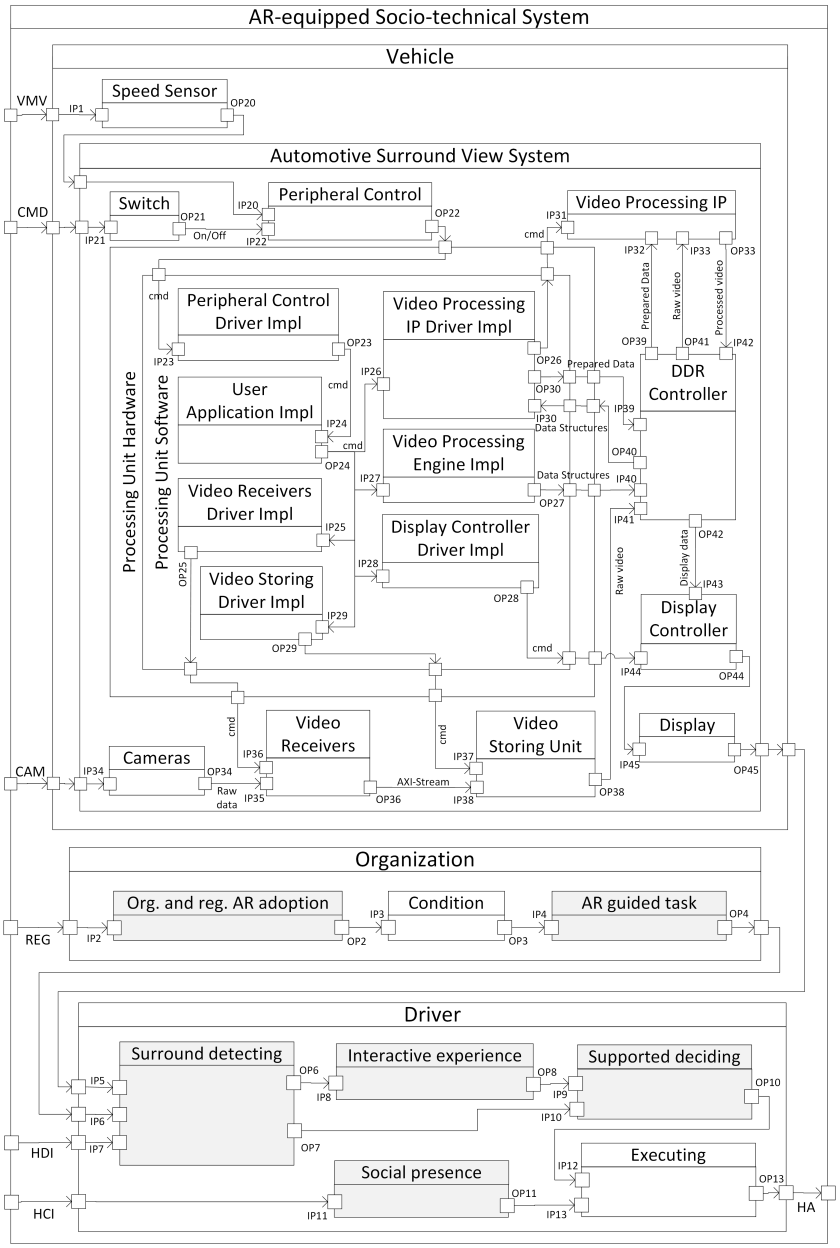

Figure 10.8: AR-equipped socio-technical system modelling

modeled in Subsection 10.3.4.

2. Second step is shown by providing FPTC rules, which are used for linking possible failure inputs of each component to failure outputs. ”IP.variable → OP.variable” shows propagational behavior of the component, which means that any failure in its input is propagated to its output. (FPTC rules for modeled sub-components are shown in Table 10.1-10.6)

3. Third step is to consider possible failures in inputs of the system to evaluate failure propagation. In this example, we inject noFailure to four inputs of the system, because we aim at analyzing system for scenarios that failure is originating from our modeled system.

4. Fourth step is calculating the failure propagations. We consider three scenarios and show the analysis results in Figure 10.9 - 10.11.

5. Last step is back propagation of results (Shown in Figure 10.12). Interpretation of the back-propagated results can be used to make decision about design change or defining safety barrier, if it is required.

Scenario 1:

- **Description of the scenario:** In this scenario, we assume that the road transport organization has not updated rules and regulations based on AR technology. Therefore, this component will produce an omission failure.

- **Modeling failure behavior:** We show the failure propagation with underlined FPTC rules, which are the rules that are activated, shown in Figure 10.9. In this scenario, surround view sub-components behave as

Table 10.1: Modeling failure behavior of components

Name of the component	Possible input failures	Possible output failures	FPTC rules (Rational)
Camera	IP34: omission, valueSubtle	OP34: omission, valueSubtle	IP34.variable → OP34.variable; (Input with failure variable, which means (omission, valueSubtle) leads to output with failure variable)
Video Receiver	IP35: late, omission, valueSubtle IP36: late, omission, commission, valueSubtle	OP36: late, omission, valueSubtle, commission	IP35.noFailure, IP36.noFailure → OP36.noFailure; IP35.variable, IP36.noFailure → OP36.variable; IP35.noFailure, IP36.variable → OP36.variable; IP35.variable, IP36.variable → OP36.variable; IP35.wildcard, IP36.omission → OP36.omission; IP35.omission, IP36.wildcard → OP36.omission; IP35.late, IP36.commission → OP36.commission; IP35.late, IP36.valueSubtle → OP36.valueSubtle; IP35.valueSubtle, IP36.late → OP36.valueSubtle; IP35.valueSubtle, IP36.commission → OP36.valueSubtle;
Video Storing Unit	IP38: late, omission, valueSubtle IP37: late, omission, commission, valueSubtle	OP38: late, omission, valueSubtle, commission	IP38.noFailure, IP37.noFailure → OP38.noFailure; IP38.variable, IP37.noFailure → OP38.variable; IP38.noFailure, IP37.variable → OP38.variable; IP38.variable, IP37.variable → OP38.variable; IP38.wildcard, IP37.omission → OP38.omission; IP38.omission, IP37.wildcard → OP38.omission; IP38.late, IP37.commission → OP38.commission; IP38.late, IP37.valueSubtle → OP38.valueSubtle; IP38.valueSubtle, IP37.late → OP38.valueSubtle; IP38.valueSubtle, IP37.commission → OP38.valueSubtle;
Display Controller	IP43: late, omission, valueSubtle IP44: late, omission, commission, valueSubtle	OP44: late, omission, valueSubtle	IP43.noFailure, IP44.noFailure → OP44.noFailure; IP43.variable, IP44.noFailure → OP44.variable; IP43.noFailure, IP44.variable → OP44.variable; IP43.variable, IP44.variable → OP44.variable; IP43.wildcard, IP44.omission → OP44.omission; IP43.omission, IP44.wildcard → OP44.omission; IP43.late, IP44.commission → OP44.commission; IP43.late, IP44.valueSubtle → OP44.valueSubtle; IP43.valueSubtle, IP44.late → OP44.valueSubtle; IP43.valueSubtle, IP44.commission → OP44.valueSubtle;

Table 10.2: Modeling failure behavior of components (Cont.)

Name of the component	Possible input failures	Possible output failures	FPTC rules
Video Processing IP	IP31: late, omission, valueSubtle IP32: late, omission, valueSubtle IP33: late, omission, valueSubtle	OP33: late, omission, valueSubtle	IP31.noFailure, IP32.noFailure, IP33.noFailure → OP33.noFailure; IP31.omission, IP32.wildcard, IP33.wildcard → OP33.omission; IP31.wildcard, IP32.omission, IP33.wildcard → OP33.omission; IP31.wildcard, IP32.wildcard, IP33.omission → OP33.omission; IP31.late, IP32.noFailure, IP33.noFailure → OP33.late; IP31.noFailure, IP32.late, IP33.noFailure → OP33.late; IP31.noFailure, IP32.noFailure, IP33.late → OP33.late; IP31.value, IP32.noFailure, IP33.noFailure → OP33.valueSubtle; IP31.noFailure, IP32.value, IP33.noFailure → OP33.valueSubtle; IP31.noFailure, IP32.noFailure, IP33.valueSubtle → OP33.valueSubtle; IP31.late, IP32.valueSubtle, IP33.noFailure → OP33.valueSubtle; IP31.valueSubtle, IP32.late, IP33.noFailure → OP33.valueSubtle; IP31.noFailure, IP32.late, IP33.valueSubtle → OP33.valueSubtle; IP31.noFailure, IP32.valueSubtle, IP33.late → OP33.valueSubtle; IP31.valueSubtle, IP32.noFailure, IP33.late → OP33.valueSubtle; IP31.late, IP32.noFailure, IP33.valueSubtle → OP33.valueSubtle; IP31.late, IP32.late, IP33.late → OP33.late; IP31.valueSubtle, IP32.valueSubtle, IP33.valueSubtle → OP33.valueSubtle; IP31.late, IP32.late, IP33.valueSubtle → OP33.valueSubtle; IP31.valueSubtle, IP32.late, IP33.late → OP33.valueSubtle; IP31.late, IP32.valueSubtle, IP33.late → OP33.valueSubtle; IP31.valueSubtle, IP32.late, IP33.valueSubtle → OP33.valueSubtle; IP31.valueSubtle, IP32.valueSubtle, IP33.late → OP33.valueSubtle; IP31.late, IP32.valueSubtle, IP33.valueSubtle → OP33.valueSubtle;

Table 10.3: Modeling failure behavior of components (Cont.)

Name of the component	Possible input failures	Possible output failures	FPTC rules
Peripheral Control Driver Imp	IP23: late, omission, commission, valueSubtle	OP23: late, omission, commission, valueSubtle	IP23.variable → OP23.variable;
User Application Imp	IP24: late, omission, commission, valueSubtle	OP24: late, omission, commission, valueSubtle	IP24.variable → OP24.variable;
Video Receiver Driver Imp	IP25: late, omission, valueSubtle, commission	OP25: late, omission, commission, valueSubtle	IP25.variable → OP25.variable;
DDR Controller	IP39: Late, omission, valueSubtle IP40: late, omission, valueSubtle IP41: late, omission, valueSubtle IP42: late, omission, valueSubtle	OP39: late, omission, valueSubtle OP40: late, omission, valueSubtle OP41: late, omission, valueSubtle OP42: late, omission, valueSubtle	IP39.variable, IP40.wildcard, IP41.wildcard, IP42.wildcard → OP39.variable; IP39.wildcard, IP40.variable, IP41.wildcard, IP42.wildcard → OP40.variable; IP39.wildcard, IP40.wildcard, IP41.variable, IP42.wildcard → OP41.variable; IP39.wildcard, IP40.wildcard, IP41.wildcard, IP42.variable → OP42.variable;
Display Controller Driver Imp	IP28: late, omission, valueSubtle, commission	OP28: late, omission, commission, valueSubtle	IP28.variable → OP28.variable;
Display	IP45: late, omission, commission, valueSubtle	OP45: late, omission, commission, valueSubtle	IP45.variable → OP45.variable;
Org. and reg. AR adoption	IP2: late, omission, valueSubtle, valueCoarse	OP2: late, omission, valueSubtle, valueCoarse	IP2.variable → OP2.variable;

Table 10.4: Modeling failure behavior of components (Cont.)

Name of the component	Possible input failures	Possible output failures	FPTC rules
Condition	IP3: late, omission, valueSubtle, valueCoarse	OP3: late, omission, valueSubtle, valueCoarse	IP3.variable → OP3.variable;
AR guided task	IP4: late, omission, valueSubtle, valueCoarse	OP4: late, omission, valueSubtle, valueCoarse	IP4.variable → OP4.variable;
Video Processing IP driver Imp	IP30: late, omission, valueSubtle IP26: late, omission, commission	OP26: late, omission, commission, valueSubtle OP30: late, omission, valueSubtle	IP26.noFailure, IP30.noFailure → OP26.noFailure, OP30.noFailure; IP26.variable, IP30.variable → OP26.variable, OP30.variable; IP30.valueSubtle, IP26.late → OP30.valueSubtle, OP26.late; IP30.wildcard, IP26.omission → OP26.omission, OP30.omission; IP30.omission, IP26.wildcard → OP30.valueSubtle, OP26.valueSubtle; IP30.late, IP26.commission → OP30.commission, OP26.valueSubtle; IP30.valueSubtle, IP26.commission → OP30.commission, OP26.valueSubtle;
Social presence	IP11: late, omission, valueSubtle	OP11: late, omission, valueSubtle	IP11.noFailure → OP11.noFailure; IP11.late → OP11.late; IP11.valueSubtle → OP11.valueSubtle; IP11.omission → OP11.omission;
Interactive experience	IP8: late, omission, valueSubtle	OP8: late, omission, valueSubtle	IP8.noFailure → OP8.noFailure; IP8.late → OP8.late; IP8.valueSubtle → OP8.valueSubtle; IP8.omission → OP8.omission;
Supported Deciding	IP9: late, omission, valueSubtle IP10: late, omission, valueSubtle	OP10: late, omission, valueSubtle	IP9.noFailure, IP10.noFailure → OP10.noFailure; IP9.variable, IP10.noFailure → OP10.variable; IP9.noFailure, IP10.variable → OP10.variable; IP9.variable, IP10.variable → OP10.variable; IP9.wildcard, IP10.omission → OP10.omission; IP9.omission, IP10.wildcard → OP10.omission; IP9.late, IP10.valueSubtle → OP10.valueSubtle; IP9.valueSubtle, IP10.late → OP10.valueSubtle;

Table 10.5: Modeling failure behavior of components (Cont.)

Name of the component	Possible input failures	Possible output failures	FPTC rules
Executing	IP12: late, omission, valueSubtle IP13: late, omission, valueSubtle	OP13: late, omission, valueCoarse	IP12.noFailure, IP13.noFailure → OP13.noFailure; IP12.late, IP13.noFailure → OP13.late; IP12.noFailure, IP13.late → OP13.late; IP12.late, IP13.late → OP13.late; IP12.valueSubtle, IP13.noFailure → OP13.valueCoarse; IP12.noFailure, IP13.valueSubtle → OP13.valueCoarse; IP12.valueSubtle, IP13.valueSubtlev → OP13.valueCoarse; IP12.wildcard, IP13.omission → OP13.omission; IP12.omission, IP13.wildcard → OP13.omission; IP12.late, IP13.valueSubtle → OP13.valueCoarse; IP12.valueSubtle, IP13.late → OP13.valueCoarse;
Switch	IP21: late, omission, commission	OP21: late, commission, omission	IP21.variable → OP21.variable;
Peripheral Control	IP20: late, omission, valueSubtle IP22: late, omission, commission, value	OP22: late, omission, commission, value	IP20.noFailure, IP22.noFailure → OP22.noFailure; IP20.variable, IP22.noFailure → OP22.variable; IP20.noFailure, IP22.variable → OP22.variable; IP20.variable, IP22.variable → OP22.variable; IP20.wildcard, IP22.omission → OP22.omission; IP20.omission, IP22.wildcard → OP22.omission; IP20.late, IP22.commission → OP22.commission; IP20.late, IP22.valueSubtle → OP22.valueSubtle; IP20.valueSubtle, IP22.late → OP22.valueSubtle; IP20.valueSubtle, IP22.commission → OP22.valueSubtle;
Video Processing Engine Imp	IP27: late, omission, valueSubtle	OP27: late, omission, valueSubtle	IP27.variable → OP27.variable;
Video Storing Driver Imp	IP29: late, omission, valueSubtle, commission	OP29: late, omission, commission, valueSubtle	IP29.variable → OP29.variable;

Table 10.6: Modeling failure behavior of components (Cont.)

Name of the component	Possible input failures	Possible output failures	FPTC rules
Surround Detecting	IP5: late, omission, valueSubtle IP6: late, omission, valueSubtle IP7: omission, valueSubtle, late	OP6: late, omission, valueSubtle OP7: late, omission, valueSubtle	IP5.noFailure, IP6.noFailure, IP7.noFailure → OP6.noFailure, OP7.noFailure; IP5.omission, IP6.wildcard, IP7.wildcard → OP6.omission, OP7.omission; IP5.wildcard, IP6.omission, IP7.wildcard → OP6.omission, OP7.omission; IP5.wildcard, IP6.wildcard, IP7.omission → OP6.omission, OP7.omission; IP5.late, IP6.noFailure, IP7.noFailure → OP6.late, OP7.late; IP5.noFailure, IP6.late, IP7.noFailure → OP6.late, OP7.late; IP5.noFailure, IP6.noFailure, IP7.late → OP6.late, OP7.late; IP5.valueSubtle, IP6.noFailure, IP7.noFailure → OP6.valueSubtle, OP7.valueSubtle; IP5.noFailure, IP6.valueSubtle, IP7.noFailure → OP6.valueSubtle, OP7.valueSubtle; IP5.noFailure, IP6.noFailure, IP7.valueSubtle → OP6.valueSubtle, OP7.valueSubtle; IP5.late, IP6.valueSubtle, IP7.noFailure → OP6.valueSubtle, OP7.valueSubtle; IP5.valueSubtle, IP6.late, IP7.noFailure → OP6.value, OP7.value; IP5.noFailure, IP6.late, IP7.valueSubtle → OP6.valueSubtle, OP7.valueSubtle; IP5.noFailure, IP6.valueSubtle, IP7.late → OP6.valueSubtle, OP7.valueSubtle; IP5.valueSubtle, IP6.noFailure, IP7.late → OP6.valueSubtle, OP7.valueSubtle; IP5.late, IP6.noFailure, IP7.valueSubtle → OP6.valueSubtle, OP7.valueSubtle; IP5.late, IP6.late, IP7.late → OP6.late, OP7.late; IP5.valueSubtle, IP6.valueSubtle, IP7.valueSubtle → OP6.valueSubtle, OP7.valueSubtle; IP5.late, IP6.late, IP7.valueSubtle → OP6.valueSubtle, OP7.valueSubtle; IP5.valueSubtle, IP6.late, IP7.late → OP6.valueSubtle, OP7.valueSubtle; IP5.late, IP6.valueSubtle, IP7.late → OP6.valueSubtle, OP7.valueSubtle; IP5.valueSubtle, IP6.late, IP7.valueSubtle → OP6.valueSubtle, OP7.valueSubtle; IP5.valueSubtle, IP6.valueSubtle, IP7.late → OP6.valueSubtle, OP7.valueSubtle; IP5.late, IP6.valueSubtle, IP7.valueSubtle → OP6.valueSubtle, OP7.valueSubtle;

propagational and propagate noFailure from inputs to output. Organization and regulation AR adoption behaves as source and while its input is noFailure, it has omission failure in its output. This activated rule is shown on this component.

- **Analysis of system behavior:** Omission failure in Organization and regulation AR adoption propagates through condition, AR guided task and in surround detecting it transforms to valueSubtle. The reason for this transformation is that omission failure in IP6 means that AR guided task is not defined by organization. This means that surround detecting would be done incorrectly because its input is not provided and this leads to valueSubtle failure in its output. ValueSubtle propagates to interactive experience and supported deciding and transforms to valueCoarse in executing. The reason for this transformation is that if there is value failure in executing function it can be detected by user, which means valueSubtle transforms to valueCoarse.
- **Interpreting the results:** Based on back propagation of the results, shown in Figure 10.12, we can explain how the rules have been triggered. ValueCoarse on OP13 is because of valueSubtle on IP12 and noFailure on OP11. ValueSubtle on IP12 is because of valueSubtle on OP10 and we continue this back propagation to reach a component originating the failure, which is component with input IP2 that is organization and regulation AR adoption. In this case, a solution would be an instruction for organization and regulation to update their rules and regulations based on AR technology. Then, the failure behavior will be updated and failure propagation analysis can be repeated to see the results.

  It is not possible to detect risks originated from failure in updating rules and regulations based on AR technology, without using the proposed representation means, because using these representation means or modeling elements provide the possibility to analyze their failure propagation and provides the possibility to analyze effect of these failures on system behavior. Then based on analysis results decision about design change or fault mitigation mechanisms would be taken.

Scenario 2:

- **Description of the scenario:** In this scenario, we assume that driver doesn't have interactive experience. Therefore, this component will

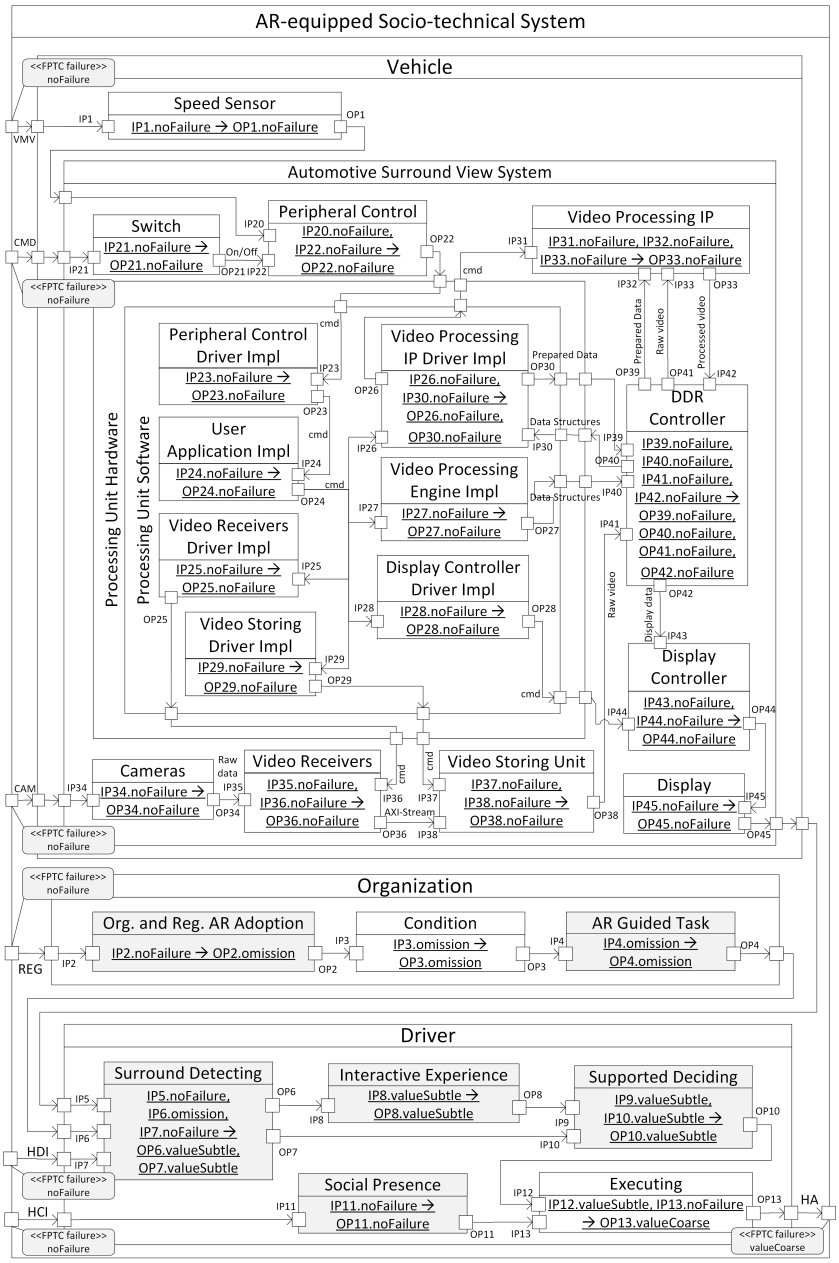

Figure 10.9: Analyzing AR-equipped socio-technical system (Scenario1)

produce a valueSubtle failure.

- **Modeling failure behavior:** We show the failure propagation with underlined FPTC rules, which are the rules that are activated, shown in Figure 10.10. Similar to the first scenario, surround view sub-components behave as propagational and propagate noFailure from inputs to output. Interactive experience behaves as source and while its input is noFailure, it has valueSubtle failure in its output. This activated rule is shown on this component.

- **Analysis of system behavior:** ValueSubtle failure in interactive experience propagates through supported deciding and in executing it transforms to valueCoarse. Similar to the first scenario, the reason for this transformation is that if there is value failure in executing function it can be detected by user, which means valueSubtle transforms to valueCoarse.

- **Interpreting the results:** Based on back propagation of the results, shown in Figure 10.12, we can explain how the rules have been triggered. ValueCoarse on OP13 is because of valueSubtle on IP2 and noFailure on OP11. ValueSubtle on IP12 is because of valueSubtle on OP10 and we continue to IP8, which is related to interactive experience component. In this case, a solution would be to suggest that the company provide a training video for all drivers at the first time of using the system. This would change the behavior type of this component from source to other types and analysis can be repeated.

  It is not possible to detect risks originated from failure in interactive experience, without using the proposed representation means, because using these representation means or modeling elements provide the possibility to analyze their failure propagation and provides the possibility to analyze effect of these failures on system behavior. Then based on analysis results decision about design change or fault mitigation mechanisms would be taken.

Scenario 3:

- **Description of the scenario:** In this scenario, we assume that AR guided task is not defined well. So this component will produce a valueSubtle failure.

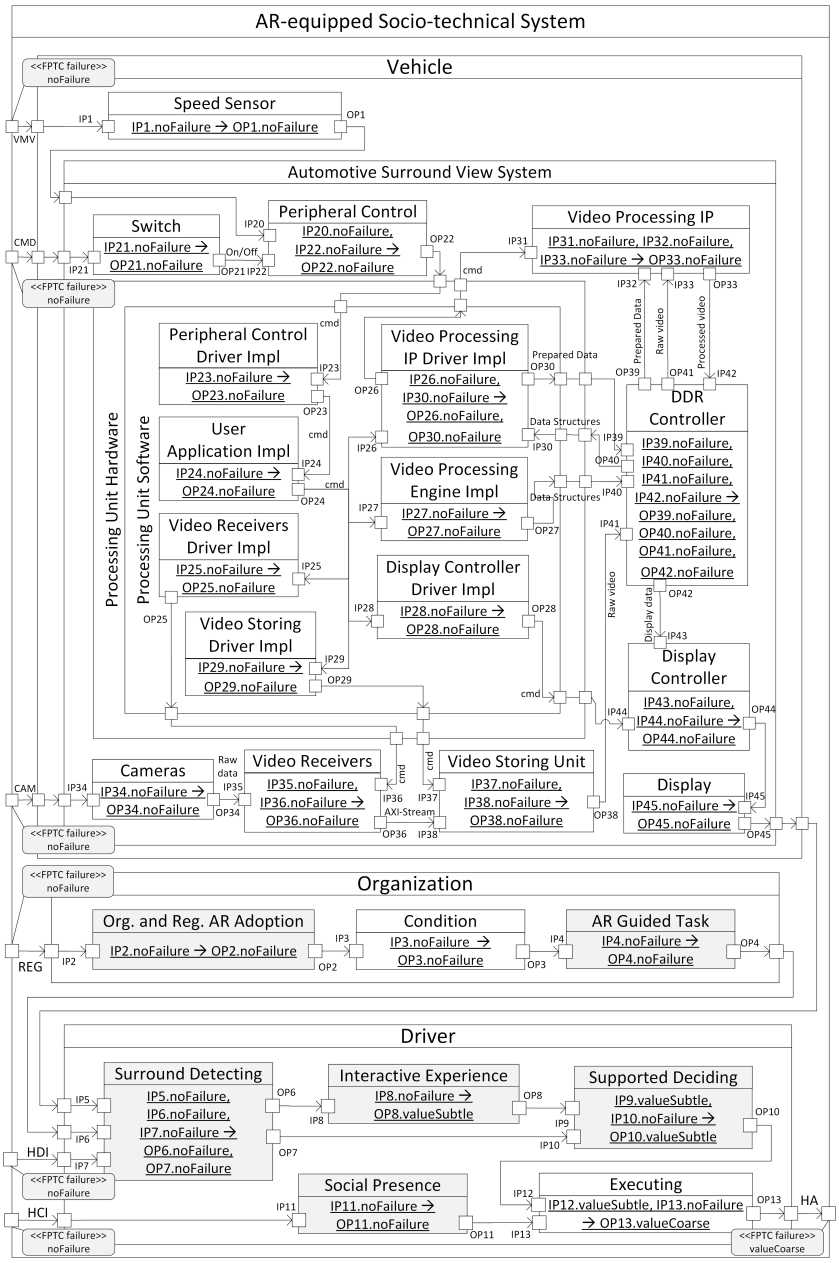

Figure 10.10: Analyzing AR-equipped socio-technical system (Scenario2)

- **Modeling failure behavior:** We show the failure propagation with underlined FPTC rules, which are the rules that are activated, shown in Figure 10.11. Similar to the previous scenarios, surround view sub-components behave as propagational and propagate noFailure from inputs to output. AR guided task behaves as source and while its input is noFailure, it has valueSubtle failure in its output. This activated rule is shown on this component.

- **Analysis of system behavior:** ValueSubtle failure in AR guided task propagates through surround detecting, interactive experience and supported deciding and in executing it transforms to valueCoarse.

- **Interpreting the results:** Based on back propagation of the results, shown in Figure 10.12, we can explain how the rules have been triggered. ValueCoarse on OP13 is originated from component with input IP4, which is AR guided task component. In this case, a solution would be to decrease the complexity of the task which AR is used for its guidance. For example dividing the task to sub-tasks decreases the complexity, which requires changes on AR design. After accomplishing the changes, modeling failure behavior should be provided to be used again in analysis.

  It is not possible to detect risks originated from failure in AR guided task, without using the proposed representation means, because using these representation means or modeling elements provide the possibility to analyze their failure propagation and provides the possibility to analyze effect of these failures on system behavior. Then based on analysis results decision about design change or fault mitigation mechanisms would be taken.

10.3.6 Lessons Learnt

In this section, we present the lessons learnt while conducting the case study. The lessons are as follows:

- **Augmented reality concepts coverage**: from a coverage point of view, as shown in Subsection 10.3.4, modeling capabilities obtained by our proposed extensions, allow architects and safety managers to model augmented reality effects on socio-technical systems by using modeling elements related to AR-extended human functions as well as modeling elements related to AR-caused faults leading to human failures. It is

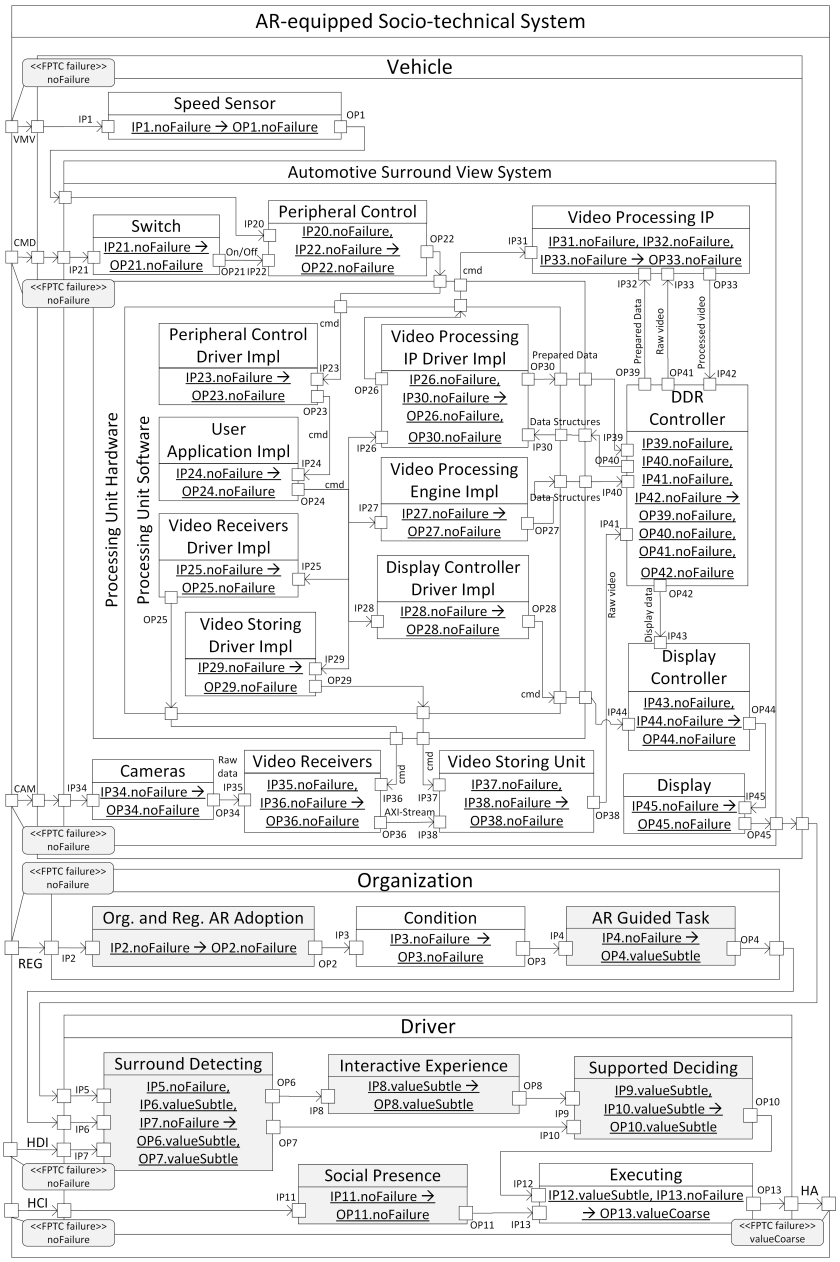

Figure 10.11: Analyzing AR-equipped socio-technical system (Scenario3)

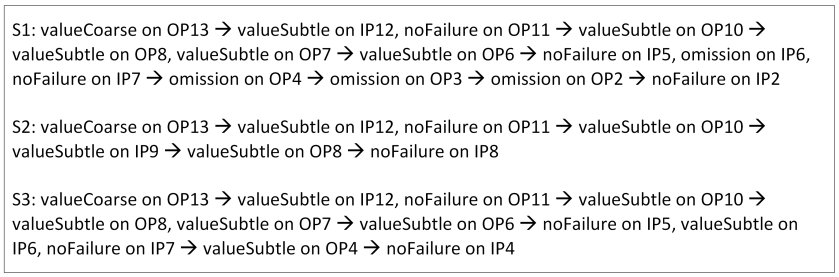

Figure 10.12: Back propagation of the results

also shown in Subsection 10.3.5 that analysis capabilities allow architects and safety managers to have at disposal means to reveal effect of AR-related dependability threats on system behavior by analyzing their failure propagation that might be effective in emerging risks within an AR-equipped socio-technical system. For example, in the first scenario, failure in updating rules and regulations based on AR technology is considered as an AR-related dependability threat and its modeling element provides representation mean for taking into account AR effect as an AR-caused fault leading to human failures. In the second scenario, failure in interactive experience and in the third scenario failure in AR guided task are considered as an AR-related dependability threats and their modeling elements provide representation means for taking into account AR effects as AR-caused faults leading to human failures.

- **Expressiveness**: Expressiveness refers to the power of a modelling language to express or describe all things required for a given purpose [30]. Set of symbols or possible statements that can be described by modelling languages can be used for measuring expressiveness. Statement means “a syntactic expression and its meaning”. As it is explained in Subsection 10.2.2, the proposed extension on human modeling elements used to extend the modeling language is based on an AR-extended human function taxonomy (AREXTax [13]), which is gained by harmonizing about 6 state-of-the-art human failure taxonomies (Norman [14], Reason [15], Rasmussen [16], HFACS (Human Factor Analysis and Classification System) [17], SERA (Systematic Error and Risk Analysis) [11], Driving [18]) and then

extending the taxonomy based on various studies and experiments on augmented reality. In addition, the proposed extension for extending organization modeling elements is based on a fault taxonomy (AREFTax [22]) containing AR-caused faults leading to human failures, which is gained by harmonizing about 5 state-of-the-art fault taxonomies (Rasmussen [16], HFACS (Human Factor Analysis and Classification System) [17], SERA (Systematic Error and Risk Analysis) [11], Driving [18] and SPAR-H (Standardized Plant Analysis Risk Human Reliability Analysis)[23]) and then extending the taxonomy based on various studies and experiments on augmented reality. According to the basis of the extensions and as it is also shown in Subsection 10.3.4, the extensions increase power of modeling language to express new AR-caused risks.

We used Concerto-FLA analysis technique as the basis of the analysis in order to disclose the advantages of our proposed AR-related extensions for CHESS framework at analysis level. Concerto-FLA uses FPTC syntax for modeling failure behavior of each component or sub-component, which includes defining FPTC rules for a component/sub-component in isolation. It is possible to define FPTC rules for the proposed AR-extended modeling elements characterizing different aspects of a component. It is important to consider possible failure modes for each input in a component/sub-component and skipping the others, because the number of FPTC rules grows exponentially with increase of the input ports. It is not conspicuous in small and academic examples, but it is really challenging if we use an industrial case study. There are also some occasions that one failure mode in input would lead to different failure modes in output. This can not be modeled using FPTC rules, because the assumption in this technique is that behavior for each component is deterministic. In industrial case studies, there would be situations with a component with non-deterministic behavior. In order to overcome this challenge, we considered the most probable situation and we modeled the component based on that situation. However, if it is required to model more complicated situations, then it is required to have more research on the extensions for techniques based on FPTC to overcome this limitation.

The generated model using our proposed AR-extended modeling elements and analysis results based on the extensions can be used as arguments based on evidences in order to provide safety case for AR-equipped industrial products to demonstrate that the system is acceptably safe to work on a given environment. However, it is required to provide also some documentation

explaining the results and how the safety requirements are achieved.

Extended human modeling elements can be used for modeling integration of human aspects with interactive systems in system testing. For example, MIODMIT architecture [31] is a generic architecture for interactive systems. As it is discussed in [32], human aspects should be considered and integrated while testing. Using extended modeling elements for modeling different aspects of human as a user of an interactive system would be of value for the system testing.

10.4 Threats to Validity

In this section, we discuss threats of validity in relation to our research based on literature [29]. Validity of a study denotes to what extent the results can be trusted.

External validity refers to possibility of generalization of the findings. We provided a case study with three scenarios from automotive domain, but the proposed extensions are not limited to specific scenarios and specific domain and the baseline for the extensions, which are AREXTax and AREFTax taxonomies are attained from taxonomies in various domains. Thus, there is the possibility of generalizing the findings for automotive domain in general and also for other domains.

Construct validity refers to the quality of choices and measurements. In our case, we used SafeConcert, which is an accepted work as the basis of our work and the proposed extensions are also based on state-of-the-art taxonomies (Norman [14], Reason [15], Rasmussen [16], HFACS (Human Factor Analysis and Classification System) [17], SERA (Systematic Error and Risk Analysis) [11], Driving [18] and SPAR-H (Standardized Plant Analysis Risk Human Reliability Analysis) [23] taxonomies) in addition to studies and experiments for the new technologies. The modeling and analysis process is done based on standardized process to increase the repeatability of the work. However, it can not be guaranteed that different people will have same answer using our proposed extensions, because it depends on the analyzer skills and ability for modeling and analysis.

Our main focus in this paper is to validate our proposed AR-related extensions for CHESS toolset on a realistic and sufficiently complex case at a level that can be found in industry. Although we were not allowed to access confidential information related to their customers, we have been able to model system architecture and failure behavior of system components using

SafeConcert metamodel, our proposed extensions and FPTC rules.

In this case study we examined the modeling and analysis capabilities of our proposed AR-related extensions through three different scenarios with different assumptions about the AR-related components' failure behavior. We have not shown that the modeling elements are complete for modelling all possible scenarios. Instead, we have focused on the provided elements to check if they are able to capture new types of system failure behaviors.

The implications of the results of the case study can not be advantageous for all different scenarios. The benefit of using our proposed extensions for a particular case depends on the ability to choose the best elements and the ability to establish failure behavior of the component related to that element. Still, this case provides evidence for the applicability and usefulness of our proposed extensions. Further investigations are required to provide more beneficial results on limitations of modelling and analysis applications.

10.5 Conclusion and Future Work

In this paper, we conducted a case study to estimate how effective our previously proposed extensions are in predicting risk caused by new AR-related threats. The extensions are for modelling and analyzing AR effects on human functioning and faults leading to human failures. We showed the analysis results by providing failure calculation manually. By implementing our proposed extensions for CHESS toolset, failure propagation calculation can be provided automatically to be used for AR-equipped socio-technical systems.

Further research is required to show the potential benefits of the proposed extensions. For example, using case studies with higher safety criticality in order to have scenarios with higher risks. In addition, having two or more teams composed of three or four experienced analysts would help to have more advanced scenarios including more complicated propagation of failures. In future, we also plan to evaluate a safety-critical socio-technical system within the rail industry, the passing of a stop signal (signal passed at danger; SPAD) [33], to verity if the results are transferrable to the rail domain.

Acknowledgment

This work is funded by EU H2020 MSC-ITN grant agreement No 764951.

Bibliography

[1] Goldiez, B.F., Saptoka, N., Aedunuthula, P.: Human performance assessments when using augmented reality for navigation. Technical report, University of Central Florida Orlando Inst for Simulation and Training (2006)

[2] Van Krevelen, D., Poelman, R.: A survey of augmented reality technologies, applications and limitations. The International Journal of Virtual Reality **9**(2) (2010) 1–20

[3] International Organization for Standardization (ISO). : ISO 26262: Road vehicles — Functional safety. (2018)

[4] Montecchi, L., Gallina, B.: SafeConcert: A metamodel for a concerted safety modeling of socio-technical systems. In: International Symposium on Model-Based Safety and Assessment, Springer (2017) 129–144

[5] Gallina, B., Sefer, E., Refsdal, A.: Towards safety risk assessment of socio-technical systems via failure logic analysis. In: 2014 IEEE International Symposium on Software Reliability Engineering Workshops, IEEE (2014) 287–292

[6] Wallace, M.: Modular architectural representation and analysis of fault propagation and transformation. Electronic Notes in Theoretical Computer Science **141**(3) (2005) 53–71

[7] Ruiz, A., Melzi, A., Kelly, T.: Systematic application of ISO 26262 on a SEooC: support by applying a systematic reuse approach. In: 2015 Design, Automation & Test in Europe Conference & Exhibition (DATE), IEEE (2015) 393–396

[8] Mazzini, S., Favaro, J.M., Puri, S., Baracchi, L.: Chess: an open source methodology and toolset for the development of critical systems. In: EduSymp/OSS4MDE@ MoDELS. (2016) 59–66

[9] Bressan, L.P., de Oliveira, A.L., Montecchi, L., Gallina, B.: A systematic process for applying the chess methodology in the creation of certifiable evidence. In: 2018 14th European Dependable Computing Conference (EDCC), IEEE (2018) 49–56

[10] CONCERTO D2.7 – Analysis and back-propagation of properties for multicore systems – Final Version: http://www.concerto-project.org/results

[11] Hendy, K.C.: A tool for human factors accident investigation, classification and risk management. Technical report, Defence Research And Development Toronto (Canada) (2003)

[12] Sheikh Bahaei, S., Gallina, B.: Towards Assessing Risk of Reality Augmented Safety-critical Socio-technical Systems. Published as proceedings annex on the International Symposium on Model-Based Safety and Assessment (IMBSA) website (2019)

[13] Sheikh Bahaei, S., Gallina, B.: Augmented reality-extended humans: towards a taxonomy of failures – focus on visual technologies. In: European Safety and Reliability Conference (ESREL), Research Publishing, Singapore (2019)

[14] Norman, D.A.: Errors in human performance. Technical report, California Univ San Diego LA JOLLA Center For Human Information Processing (1980)

[15] Reason, J.: The human contribution: unsafe acts, accidents and heroic recoveries. CRC Press (2017)

[16] Rasmussen, J.: Human errors. a taxonomy for describing human malfunction in industrial installations. Journal of occupational accidents **4**(2-4) (1982) 311–333

[17] Shappell, S.A., Wiegmann, D.A.: The human factors analysis and classification system–HFACS. Technical report, Civil Aeromedical Institute (2000)

[18] Stanton, N.A., Salmon, P.M.: Human error taxonomies applied to driving: A generic driver error taxonomy and its implications for intelligent transport systems. Safety Science **47**(2) (2009) 227–237

[19] Fu, W.T., Gasper, J., Kim, S.W.: Effects of an in-car augmented reality system on improving safety of younger and older drivers. In: 2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), IEEE (2013) 59–66

[20] Schall Jr, M.C., Rusch, M.L., Lee, J.D., Dawson, J.D., Thomas, G., Aksan, N., Rizzo, M.: Augmented reality cues and elderly driver hazard perception. Human factors **55**(3) (2013) 643–658

[21] Sheikh Bahaei, S., Gallina, B.: Extending safeconcert for modelling augmented reality-equipped socio-technical systems. In: International Conference on System Reliability and Safety (ICSRS), IEEE (2019)

[22] Sheikh Bahaei, S., Gallina, B., Laumann, K., Rasmussen Skogstad, M.: Effect of augmented reality on faults leading to human failures in socio-technical systems. In: International Conference on System Reliability and Safety (ICSRS), IEEE (2019)

[23] Gertman, D., Blackman, H., Marble, J., Byers, J., Smith, C., et al.: The SPAR-H human reliability analysis method. US Nuclear Regulatory Commission **230** (2005)

[24] Miller, M.R., Jun, H., Herrera, F., Villa, J.Y., Welch, G., Bailenson, J.N.: Social interaction in augmented reality. PloS one **14**(5) (2019) e0216290

[25] Šljivo, I., Gallina, B., Carlson, J., Hansson, H., Puri, S.: A method to generate reusable safety case argument-fragments from compositional safety analysis. Journal of Systems and Software **131** (2017) 570–590

[26] Sljivo, I., Gallina, B., Carlson, J., Hansson, H., et al.: Using safety contracts to guide the integration of reusable safety elements within iso 26262. In: 2015 IEEE 21st Pacific Rim International Symposium on Dependable Computing (PRDC), IEEE (2015) 129–138

[27] Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles: https://www.sae.org/standards/content/j3016_201806/ (2018)

[28] Dimitrakopoulos, G., Uden, L., Varlamis, I.: The Future of Intelligent Transport Systems. Elsevier (2020)

[29] Runeson, P., Höst, M.: Guidelines for conducting and reporting case study research in software engineering. Empirical software engineering **14**(2) (2009) 131

[30] Patig, S.: Measuring expressiveness in conceptual modeling. In: International Conference on Advanced Information Systems Engineering, Springer (2004) 127–141

[31] Cockton, G., Woolrych, A.: Understanding inspection methods: lessons from an assessment of heuristic evaluation. In: People and computers XV—Interaction without frontiers. Springer (2001) 171–191

[32] Canny, A., Bouzekri, E., Martinie, C., Palanque, P.: Rationalizing the need of architecture-driven testing of interactive systems. In: International Conference on Human-Centred Software Engineering, Springer (2018) 164–186

[33] Naweed, A., Trigg, J., Cloete, S., Allan, P., Bentley, T.: Throwing good money after spad? exploring the cost of signal passed at danger (spad) incidents to australasian rail organisations. Safety science **109** (2018) 157–164