Functional safety analysis of SAE conditional driving automation in time-critical situations and proposals for its feasibility

Takayuki Hirose^{1,2} · Tetsuo Sawaragi¹ · Hideki Nomoto² · Yasutaka Michiura²

Received: 30 April 2020 / Accepted: 22 September 2020 / Published online: 10 October 2020
© The Author(s) 2020

Abstract

The autonomous driving technology is rapidly developed for commercial use, aiming at the conditional driving automation or the third level of driving automation (LoDA 3). One of the most critical challenges for the achievement is the smooth authority transfer from the system to human drivers in an emergency. However, it is still on the way to find out fundamental solutions. The difficulty is closely related to the envisioned world problem, for which the concept of functional modeling could be a solution. This paper presents a safety analysis of the authority transfer problem based on the ideas of functional modeling. We discuss the safety of the authority transfer in a time-critical situation by using a simulator based on the functional resonance analysis method (FRAM). The result shows that the involvement of human drivers in driving activities is still essential even during autonomous driving. We show that the current LoDA 3 is a myth that the human drivers are supposed to be required just in non-normal situations, while they can be free from dynamic driving tasks (DDTs) in usual cases. Based on the result, this work makes some proposals for successful autonomous driving, effective human–machine collaboration, and the right design of artifacts.

Keywords Autonomous driving · Conditional driving automation · Levels of driving automation (LoDA) · Envisioned world problem · Functional resonance analysis method (FRAM)

1 Introduction

The technology of autonomous driving is rapidly developed for commercial use, aiming at conditional driving automation or the third level of driving automation (LoDA 3) issued by SAE (2016). At the LoDA 3, while the system is basically responsible for dynamic driving tasks (DDTs), human drivers also need to intervene in the operation in an emergency. One of the most critical issues for the achievement is how to

smoothly transfer the driving authority from the automated system to human drivers therein. Many researchers from both academic and industrial fields have been working on this issue. However, it is still on the way. More specifically, while the nature of this problem highly depends on specific contexts (Inagaki and Sheridan 2019; Eriksson and Stanton 2017), most of the current approaches are restricted only to identify static or even fixed time requirements to issue the take over requests (TOR).

The difficulty is closely related to what is called the envisioned world problem (Woods and Dekker 2000; Woods and Christoffersen 2002). Woods and Dekker (2000) pointed out that it is quite challenging to envision actual human–machine interactions in research and development (R&D) processes, since many factors generally interfere with the work domain, making them so complicated. The challenge of envisioning the actual work domain often results in too little information for the R&D processes at too great a cost, i.e., the nature of the envisioned world problem.

Roth and Mumaw (1995) and Woods and Dekker (2000) discussed that the functional modeling methodologies are possible solutions to overcome the problem. It enables us to

✉ Takayuki Hirose
hirose.takayuki.27v@kyoto-u.jp

Tetsuo Sawaragi
sawaragi.tetsuo.3r@kyoto-u.ac.jp

Hideki Nomoto
nomoto.hideki@jamss.co.jp

Yasutaka Michiura
michiura.yasutaka@jamss.co.jp

¹ Department of Mechanical Engineering and Science, Kyoto University, Kyoto 615-8540, Japan

² IV&V Research Laboratory, Japan Manned Space Systems Corporation (JAMSS), Tokyo 100-0004, Japan

represent human–machine interactions as a relationship of functions consisting of the interactions and investigate their nature to obtain generic insights. In this context, functional resonance analysis method (FRAM) (Hollnagel 2004, 2012) especially attracts its attention since it enables us to consider the “non-linear” interaction of human, machine, and their surrounding environment.

The objective of this research is to investigate the authority transfer problem based on the ideas of the envisioned world problem and functional modeling. In this paper, a simulation model based on FRAM (Hirose and Sawaragi 2019) is adopted for it, with which we simulate the safety of the authority transfer activity in a time-critical situation. Also, the feasibility of the current concept of the LoDA 3 is discussed based on the result. Finally, we conclude this paper with discussions of proposals for successful autonomous driving, effective human–machine collaboration, and the right design of artifacts.

2 Research and development of autonomous driving technology and its difficulties

2.1 Research and development of autonomous driving technology

We can classify the autonomous driving technologies developed so far for commercial use into one of the five levels of driving automation (LoDA) issued by SAE (2016). According to the definition, the system is basically responsible for the dynamic driving tasks (DDTs) above the LoDA 3. The research and development (R&D) of the technology currently aims at the LoDA 3 or conditional driving automation, where human drivers still have to intervene in the DDTs in an emergency.

The authority transfer from the autonomous driving system to the drivers in an emergency has primarily been one of the most challenging issues in this context. Many researchers in both academic and industrial fields have been working on the subject. Their primary interests are on how many seconds are required to request the drivers to take over the authority and disengage the automation. However, finding the fundamental solution is still on the way; their proposals often sound so ad hoc that they can be valid to just limited situations.

In this regard, Inagaki and Sheridan (2019) and Eriksson and Stanton (2017) pointed out the context dependency of conditions to transfer the driving authority. Inagaki (2003) pointed out that the conventional discussions about function allocation between human operators just focus on “who does what,” and the perspective of “who does what and WHEN” is missing there; the function allocation should

be adaptive to the situations where operational functions should be shared or traded between the human operators and systems, depending on a specific context. Eriksson and Stanton (2017) reviewed 25 kinds of researches about the responding time to the take over request (TOR) to point out that the previous researches put too much focus on the average time to issue the TOR and take over the authority. They also concluded that the mean or median values do not tell the entire story of the authority transfer through their additional experiment, and the strategy of the authority transfer should be adaptive to situations.

Despite the nature of this problem, the current context of the R&D for the autonomous driving technology is eager to find out the static or even “fixed” time for the drivers to smoothly take over the authority in an emergency. This inconsistency is now leading us to the never-ending exploitation and away from the fundamental solutions. The problem is that it is quite difficult to properly envision actual situations and practical requirements in the R&D process.

2.2 Envisioned world problem: difficulties in predicting evolving field of practice

According to Woods and Dekker (2000), the introduction of new technology is not a simple substitution of machines for people, but an intervention into the ongoing field of the practice. In this context, we cannot handle new technology as not only an object, but also a hypothesis about how technological change will transform practice and shape how people adapt, i.e., the evolution of the field of practice. It is, therefore, essential to anticipate the impact of the new technology involving human factors, and it should, in theory, be addressed in the early phase of the R&D process.

However, it is not the case in reality. The human factors are often pushed to the tail of the R&D process, because it is quite challenging to envision the lucid visions of the evolving field of practice emerging out of the “non-linear” interactions among human, machines, and their surrounding environment. Besides, empirical tests in this phase generally provide too little information at too high a cost. This is known as the envisioned world problem (Woods and Dekker 2000; Woods and Christoffersen 2002), which has been causing various issues related to the new technology. The autonomous driving technology is also the case. One of the most considerable demands for the impact of the new coming technology is to develop innovative methodologies to overcome this problem.

Cognitive task analysis is one of the possible approaches for the solution. According to Roth and Mumaw (1995), traditional methods of task analysis are to structure analyses of operator information, control requirements around specific event sequences, and preplanned response strategies for handling those event sequences. The traditional task

analysis approaches are, therefore, useful in ensuring the human–machine interfaces (HMI) support of human operators in pre-analyzed situations. On the other hand, they do not provide a principled way to identify information and control requirements to support operator performance in unanticipated situations. The methodologies of cognitive task analysis have been developed in this context.

The typical approach of the cognitive task analysis is to represent a target system as a relationship among functions of humans and machines. Abstraction hierarchy (AH) by Rasmussen (1986) or multilevel flow modeling (MFM) by Lind (2011) is one of the most well-known methodologies for it. Also, functional resonance analysis method (FRAM) (Hollnagel 2004, 2012) is now attracting its attention as an innovative method because it additionally focuses on the “non-linear” interaction of human, machine, and their surrounding environment. They do contribute to envisioning the evolving field of practice and provide generic insights for the R&D of new technology.

The current R&D of the autonomous driving technology is conducted within the framework of the envisioned world problem. It is, therefore, necessary to leave the physical or real-oriented approaches once and carry out investigations at a higher-functional abstraction level to overcome this problem; functional resonance analysis method (FRAM) is a possible solution for it.

Besides, there have actually been a number of methodologies developed for safety analysis of human–machine systems such as fault tree analysis (FTA) (Vesely et al. 1981; Lee et al. 1985; Stamatelatos et al. 2002), technique for human error-rate prediction (THERP) (Swain and Guttman 1983), Petri Net (Petri 1966; Brauer and Reisig 2009), structured analysis and design technique (SADT) (Ross 1977), or function analysis system technique (FAST) (Bytheway 2007), for example. Although they are one of the most well-known approaches of safety analysis and have been applied to many fields, their limitations have also been pointed out so far. The FTA and its successor model THERP are typical sequential methodologies of reliability analysis; the limitation is that their approach is so simple and linear that it is difficult to address complex aspects of the safety such as cognitive behavior of human beings and “non-linear” interactions among the systems’ components. The SADT and FAST are good at describing functional relationships within the systems as a diagram, but their expressiveness (i.e., semantics of functions) is still limited compared to FRAM. The Petri Net is also a representative approach of discrete event systems, but it is also limited to deal with the “non-linear” aspects of the safety as well as other approaches.

FRAM is characteristic compared to these well-known methods in terms of the following aspects. First, the scope of FRAM is not limited to failures or “why things go wrong”; its aim is rather to find out “why things go right”

and facilitate the safety based on the findings as its fundamental concepts, i.e., Safety-II and Resilience Engineering (Hollnagel 2012, 2017). Also, FRAM is not to investigate just “snapshots” of safety or systems’ state based on linear causality, contrary to the traditional approaches; it has a potential to envision dynamic characteristics of the safety or systems’ state emerging out of “non-linear” interactions among components of the target systems. Moreover, such emergent behavior essentially involves abstract, ambiguous, and qualitative aspects, for which the balance between qualitative and quantitative perspective must be taken into account; a previous study (Hirose and Sawaragi 2020) has suggested in this respect that FRAM could provide qualitative comprehensions of the emergent behavior on the basis of quantitative criteria. FRAM is different from other traditional approaches and adequate for investigating the safety of complex human–machine interactions in these respects; the overview is reviewed in the following.

2.3 Functional resonance analysis method (FRAM)

FRAM is a method to investigate the safety of systems whose operation involves human, machine, and their surrounding environment. Those systems are generally called socio-technical systems, and the operation of autonomous driving can be described in this framework as well. FRAM enables us to build functional models of the socio-technical systems and investigate their safety emerging out of the “non-linear” interaction among their elements.

FRAM explicitly puts focuses on the interaction of variabilities existing in the functions of human–machine interactions and their surrounding environment. There generally exist variabilities in the surrounding environment, caused by temporal conditions such as available resources or social demands. They interfere with the functions of human–machine interactions and induce the functional variabilities coping with the exogenous variabilities, as shown in Fig. 1; Hollnagel (2004) and Hollnagel (2012) refer to this as the efficiency and thoroughness trade-offs (ETTO) principle. All of the variabilities eventually interact with each other. Moreover, their interaction could result in unexpected outcomes, as if they were resonated, in a specific context.

Each of the function in FRAM is defined with six aspects of input, output, precondition, resource, control, and time. They enable us to describe the complex dependency among functions and build a functional model of human–machine interactions as a network of functions, which is referred to as an instance. Based on this, Fig. 1 can consequently be updated, as shown in Fig. 2; this is the basic structure of FRAM, in which the effect of the variability changes the state of the functions and even their dependency structures.

Fig. 1 Schematic illustration of socio-technical systems

Fig. 2 Schematic illustration of FRAM: functional model of socio-technical systems

3 Safety simulation of conditional driving automation in time-critical situations

3.1 Overview and objective

This simulation is to investigate the safety of the authority transfer problem in time-critical situations. The crucial difference from the traditional approaches is that this simulation does not consider the real, physical conditions such as a distance to an obstacle (e.g., construction site) ahead of the vehicle. Besides, it does not deal with the exact time needed before the issue of TOR informing when the automation will be disengaged. Instead, this case study is to model the functional safety of the operation of the autonomous driving system (ADS) and investigate how its safety is resilient against the variabilities to address the envisioned world problem.

Especially, the objective of this simulation is to investigate the feasibility of conditional driving automation at

LoDA 3, where the human driver should still be the final authority of the driving and must take evasive actions in an emergency. In such situations, the driver is required to process good situation awareness (SA) and make consequent decision making (DM) when the TOR is issued, and the ADS inevitably has to provide the appropriate cognitive support with the human driver. According to Endsley (1995), the SA consists of the three levels shown in Table 1. The formal definition of SA is broken down into three segments: perception of the elements in the environment, comprehension of the situation, and projection of future status. To make a smooth authority transfer from automation to a human driver, the systems should also guarantee supports that are appropriate for each of the three segments of a driver's SA. That is, a human driver, for whom the TOR is issued and prepares for the authority transfer from automation, has to attain successful SA conducted by the support of the system. In an emergency, it would be ideal for supporting all segments of SA thoroughly as earlier as possible, but it would not always be possible. Its success would depend upon which cognitive status a driver is and upon what support is provided by the system. Complex behaviors may emerge, and the functional simulation is needed to find out the best human-automation collaboration.

3.2 Methodology

The simulation is carried out employing the FRAM model developed by Hirose and Sawaragi (2019). According to Hollnagel (2012), FRAM is a method rather than a model. This means that the original FRAM just provides the concept of how to describe and understand the safety of socio-technical systems as described in the previous section, and it has been necessary to build a specific model for its practical use. Specifically, the definitions of essential entities of FRAM—such as variabilities and their interactions—are so conceptual that many people are still eager to find out how to use it. In response to this need, they have developed a simulation model of FRAM whose architecture is consistent with Fig. 2. The feature of their model is that the variabilities of socio-technical systems are defined based on Fuzzy

Table 1 Levels of SA

Level	Description
1. Perception	To detect some data or elements of what is going on
2. Comprehension	To integrate the elements and comprehend the current situation or the reason why it is happening
3. Projection	To anticipate future status based on the comprehension of the current situation, leading to DM

CREAM—an extended model of cognitive reliability and error analysis method (CREAM) (Hollnagel 1998), and their interactions among those variabilities are formulated as well.

According to CREAM, the surrounding factors creating the context described in Figs. 1 and 2 can collectively be defined as 11 common performance conditions (CPCs) shown in Table 2; it should be noted that the number of CPCs were originally 9 (Hollnagel 1998), and 2 more CPCs: “Available resources” and “Quality of communication” were added later as a result of further research (Hollnagel 2004). They are generally evaluated with linguistic values such as “Available resource is enough” or “Number of simultaneous goals are too much,” and their evaluations result in identification of another index called “control mode.”

The control mode originates from contextual control model (COCOM) (Hollnagel 2003). In COCOM, the control is described in terms of how we do things and how actions are chosen and executed according to their surrounding environment. Moreover, the outcome of the control prescribes a certain sequence of the possible actions, and the sequence is constructed by the environment rather than pre-defined. The control can obviously occur on several levels or in several modes, and that is why four linguistic values shown in Table 3 have been introduced as the control mode.

In this context, Fuzzy CREAM enables us to consider quantitative and continuous aspects of CREAM. That is, membership functions of the linguistic CPC states are associated with continuous scores ranging from 0 to 100, and the

fuzzy reasoning based on their evaluations provides a control mode as a crisp value called probability of action failure (PAF); the higher CPC scores represents the better state of each CPC. Several studies proposed Fuzzy CREAM methodologies (Konstandinidou et al. 2006; Yang et al. 2013; Ung 2015), among which “weighted CREAM model (Ung 2015)” was adopted for the FRAM model since its process is systematically well organized, and the impact or weight of each CPC on control mode is also taken into account.

Based on the idea of Fuzzy CREAM, the proposed FRAM model assumes that a state in each function can be described in terms of the control mode, and the state can change according to the state of their surrounding environment created by CPCs and their weight; the variabilities in each function and the surrounding environment are therefore defined as the change of their continuous control mode (crisp value of PAF) and CPC scores, respectively. These definitions are based on the fact that the origin of the FRAM functions is derived from the framework of work domain analysis (WDA) (Rasmussen et al. 1994; Vicente 1999). The WDA is a framework to identify the functional properties of a system at different levels of abstraction, and represent them in a means–end hierarchy to show the relationships among functions in each abstraction level. In this framework, the functions are described as predicates and essentially qualitative properties. The functional safety in this context refers to the safety of those functions and is therefore described as qualitative properties as well. That is why the control mode can be introduced to represent such qualitative properties, and moreover, the representation, i.e., control mode is now supported by semi-quantitative criteria based on Fuzzy CREAM as well; those definitions of variabilities consequently therefore enable us to investigate the functional safety on the basis of the balance between qualitative and quantitative perspectives.

The numerical definition of the variabilities enables us to formulate the interaction of variabilities, and the simulation model eventually repeats a set of the following processes: (1) variabilities of CPC scores induce that of control mode in each function based on Fuzzy CREAM; (2) the variabilities of functions propagate to their downstream functions and interact with each other according to one of the formulations; (3) the effect of interaction is looped back and changes the scores of surrounding CPCs again according to another

Table 2 List of CPCs (Hollnagel 1998, 2004)

ID No.	CPC
1	Available resources
2	Adequacy of training and experience
3	Quality of communication
4	Adequacy of human–machine interface (HMI)
5	Availability of procedures
6	Working conditions
7	Number of simultaneous goals
8	Available time
9	Circadian rhythm
10	Crew collaboration quality
11	Adequacy of organization

Table 3 Description of control mode (Hollnagel 1998, 2003)

Control mode	Each control mode represents a situations in which:
Strategic	Tasks can be carried out based on sophisticated strategies and enough comprehension of global context
Tactical	People can carry out their tasks based on planning and following procedures, but their scope is limited
Opportunistic	The next action heavily relies on the salient features of a current context rather than stable strategies
Scrambled	People are too upset to choose a proper action; an extreme case here would be the state of momentary panic

formulation. In the end, the change of control mode in each function with respect to the simulation time: $T[-]$ can be obtained as a result of the simulation.

Here, it should be noted that the dimensionless simulation time: $T[-]$ is defined as the number of loops in which a set of the processes is repeated and different from the real time. The concept of simulation time is rather consistent with that adopted in qualitative reasoning (Kuipers 1986, 1994, 2001), in which the time points are defined as those points in time when a qualitative state of the simulation model (i.e., the qualitative value of any variable) changes. Therefore, the time in the simulation world goes on independent of that in the real world; the very long time in the simulation world can correspond to a moment in the real world, and vice versa. The simulation time: $T[-]$ advances based on this idea as well, and the qualitative change is brought about by the loop of above three processes.

3.3 Initial setting of simulation

Some parameters should be set to start the simulation. The first item of the initial setting is to model the target system with the functions. The second item is to set the weight of CPCs for each function. We also need to come up with a simulation scenario and describe it with parameters of the FRAM model. The details are shown below.

3.3.1 Functions and their potential couplings

The operation of autonomous driving is modeled as shown in Fig. 3 in this simulation. The model was built based on fundamental tasks of the drivers, including SA, DM, and their manual operations. Further details about each

function are described below. Note that the potential couplings are currently illustrated with dotted lines since they just represent the possible dependency among functions. The potential couplings will be instantiated depending on a specific context or situation later on.

The critical goal of a driver is to drive without any troubles, and it is represented as a generic function: To DRIVE. This function generates an actual driving context or situation as its output, and both humans and machines initially detect it. This triggers the establishment of a driver's SA, and additional functions subsequently represent this SA process.

The initial detection process of SA is represented with the following three additional functions: To PAY ATTENTION TO TRAFFIC CONDITIONS; To PAY ATTENTION TO IN-CAR DISPLAY; To SENSE DRIVING INFORMATION by the system. The input of these functions are potentially coupled with the output of To DRIVE. Moreover, the first two functions with their couplings are related to the establishment of the SA Lv. 1.

The establishment of the SA Lv. 1 leads to the development of the SA Lv. 2. Thus, an additional function: To COMPREHEND CURRENT DRIVING SITUATION is defined. Also, this function is coupled with the upstream functions of To PAY ATTENTION TO TRAFFIC CONDITIONS and To PAY ATTENTION TO IN-CAR DISPLAY.

Further, the SA Lv. 3 is developed based on the establishment of the SA Lv. 2, which is represented by the additional function of To PLAN AND IDENTIFY NEXT ACTIONS. The input of this function is connected with the output of To COMPREHEND CURRENT DRIVING SITUATION, and the output provides the driver's decision whether to continue the autonomous driving or not. In the end, the decision provides the inputs of additional two functions: To CONTINUE AUTONOMOUS DRIVING

Fig. 3 Potential couplings among FRAM functions: operation of ADS

and TO TAKE EVASIVE ACTIONS, whose outputs are fed back to the input of the generic function: TO DRIVE.

Besides the above structure of this model, the output of TO SENSE DRIVING INFORMATION is also coupled with the input of TO ALERT DRIVERS TO PREPARE FOR EMERGENCY. This makes the downstream function ready for launching alerts. Moreover, the output of the alerting function is potentially coupled with the inputs of the functions (the functions of 2, 5, 6, and 9) related to each level of the driver's SA.

By the way, the instantiation of potential couplings shown in Fig. 3 generates a large number of structures of FRAM model which is called instances; the number of possible instances indeed amount for $2^{19} = 524,288$ patterns in this case since there are 19 dotted lines, i.e., potential couplings in the general structure of FRAM model. However, it is apparently irrational to consider all of them, and the possible instantiations should therefore be limited on the basis of some constraints in advance. In this respect, the general structure of FRAM model in Fig. 3 can be reconfigured based on the perceptual cycle (Neisser 1976; Smith and Hancock 1995), and the constraints on the possible instantiations can be obtained based on its analysis.

The perceptual cycle describes a cyclic process in which cognitive agents (e.g., human beings or automation) take some actions directed by their internally held knowledge—or schemata of the world, and their outcome modifies the original knowledge or schemata as shown in Fig. 4. Specifically, the agents have a schema of present environment as a part of cognitive map of the world, and it directs a part of locomotion and action for perceptual exploration of the world. This exploration samples available information of actual present environment, and the sampled information updates the original schema of present environment. This

Fig. 4 Perceptual cycle (Neisser 1976; Smith and Hancock 1995)

cyclic relationship is invariant and must exist for any reasons. Based on this framework, the FRAM model in Fig. 3 can be redrawn as shown in Fig. 5.

Figure 5 illustrates another configuration of Fig. 3. In this configuration, each function is distributed and grouped into one of three elements of the perceptual cycle: TO CONTINUE AUTONOMOUS DRIVING and TO TAKE EVASIVE ACTIONS were grouped into the “Locomotion and action” of perceptual cycle; TO DRIVE was distributed to the “Actual world (potentially available information)” for convenience because the output of this function was originally defined to provide actual driving context or situation, rather than some specific actions as its output; the rest of functions were grouped into cognitive part, i.e., “Cognitive map of the world” of the perceptual cycle. Here, only the position of each function was changed, and other parameters such as potential couplings among functions are not modified at all. In addition, the grouped functions with potential couplings, i.e., dotted lines are currently corresponding to the entire part of each triangle in Fig. 5; the functions get to represent the part of triangles clipped by the circle, i.e., the area with solid lines in Fig. 5 as a result of instantiations.

A topological analysis of this reconfigured model provides at least three constraints on the instantiation. The first constraint is that dependency of functions interconnecting the groups, i.e., “Locomotion and action,” “Actual world,” and “Cognitive map of the world” of perceptual cycle must exist due to the invariance of their relationships. In other words, the potential couplings or functions forming this dependency basically have to be instantiated. The second constraint is that the two functions in the “Locomotion and action” group cannot be instantiated at the same time; if the function: TO CONTINUE AUTONOMOUS DRIVING is instantiated and active, the other function: TO TAKE EVASIVE ACTIONS should be deactivated. It is therefore allowed us to instantiate only potential couplings or functions belonging to the “knowledge” group with some freedom. In the end, the third constraint is that functions or potential couplings that we can arbitrary instantiate is limited to TO PAY ATTENTION TO TRAFFIC CONDITIONS, TO PAY ATTENTION TO IN-CAR DISPLAY, or the outputs of TO ALERT DRIVERS TO PREPARE FOR EMERGENCY. This is because the rest of functions in this group are related to internal process of humans or machines, and their intentional instantiation could result in too much subjectivity/arbitrariness of the simulation. These constraints significantly limit the number of possible instantiations, and their actual patterns are provided in the following part of simulation scenario construction.

3.3.2 Parameters of CPCs

CPC weight is the sensitivity of a function against each CPC. We define the CPC weights by using a paired

Fig. 5 Another configuration of FRAM model based on perceptual cycle

comparison process, which is a part of the analytic hierarchy process (AHP) (Saaty 1990). This is because the CPC weight is based on qualitative relationships between a FRAM function and its related CPCs, and it might involve too much subjectivity if it were defined intuitively. Hirose and Sawaragi (2020) have already adopted this, and the CPC weight is calculated in the same way here. The CPC weight in each function is consequently identified as shown in Table 4; the identification numbers 1–9 after the second column correspond to that of the functions shown in Fig. 3.

3.3.3 Simulation scenario

Simulation scenarios are prepared to envision the emerging functional behaviors, or how an event of the authority transfer may affect the driver's following cognitive activities exhaustively. The scenarios consist of the corresponding settings of multiple parameters such as CPC scores and potential couplings (i.e., instantiations) intervening into ongoing simulation processes. The manual settings of these parameters can be regarded as additional variabilities, and they were classified into two categories in this simulation: the

Table 4 CPC weight in each function

List of CPCs	Identification no. of functions shown in Fig. 3								
	1	2	3	4	5	6	7	8	9
Available resources	0.102	0.240	0.0746	0.204	0.138	0.179	0.173	0.191	0.1600
Adequacy of training and experience	0.102	0.0932	0.0746	0.0889	0.138	0.0640	0.0324	0.0213	0.0573
Quality of communication	0.0355	0.0333	0.0284	0.0449	0.0462	0.0234	0.0324	0.0638	0.0573
Adequacy of human–machine interface (HMI)	0.102	0.0333	0.0746	0.0188	0.138	0.0640	0.0324	0.0638	0.1600
Availability of procedures	0.0355	0.0333	0.0284	0.0188	0.0154	0.0234	0.0324	0.0213	0.0214
Working conditions	0.102	0.0932	0.200	0.204	0.138	0.179	0.214	0.191	0.1600
Number of simultaneous goals	0.102	0.0932	0.200	0.0889	0.138	0.179	0.173	0.191	0.1600
Available time	0.269	0.240	0.200	0.204	0.138	0.179	0.214	0.191	0.1600
Circadian rhythm	0.102	0.0932	0.0746	0.0449	0.0462	0.0640	0.0324	0.0213	0.0214
Crew collaboration quality	0.0355	0.0333	0.0284	0.0628	0.0462	0.0234	0.0324	0.0213	0.0214
Adequacy of organization	0.0151	0.0149	0.0153	0.0188	0.0154	0.0213	0.0324	0.0213	0.0191

one is called exogenous variabilities causing the time-critical situations, triggered by the manual change of CPC scores; the other is called endogenous variabilities that are structural changes, i.e., instantiations of FRAM model in response to the exogenous variabilities. The simulation scenarios were constructed based on these two kinds of variabilities so that the exogenous variabilities cause a time-critical situation and induce the endogenous variabilities. That is, the following scenarios are to investigate the composite effect of not only the exogenous variabilities caused by the change of CPC scores, but also the endogenous variabilities that reconstruct the FRAM model during the ongoing simulation process.

The first step of the scenario construction is to prepare a verbal description of the scenario as follows:

Verbal description: A car is driving with its ADS, and the driver is currently distracted by some activities such as reading a book or using a smartphone rather than being engaged in the driving tasks, leaving them

to the automation. Everything is going well at first. After a while, the system suddenly launches an alert demanding the driver to take over the authority of driving immediately. The driver must respond to this and take some evasive actions.

This verbal description can be represented as a sequential diagram in Fig. 6. The following steps of the scenario construction are to define several events described in the diagram, i.e., Initial state, Variability-triggering event (exogenous variability), variations of Instantiation 1s, and Instantiation 2 (endogenous variabilities), as the manual operation of simulation parameters.

First of all, the Initial state is set as following, representing the situation until the driver and ADS face a time-critical situation:

Initial state: all CPC scores are equally set to 100, and the partial structure of Fig. 3 is instantiated, as shown in Fig. 7. Therein, two functions related to the SA Lv.

Fig. 6 Sequential diagram of verbal description

Fig. 7 A prototype of instantiations after the arrival of the alert: it will be common in the scenarios shown in Table 5

1 are temporally missing in the instance; this instance represents the active functional components existing during the autonomous driving without any involvement of the driver.

The simulation is then triggered by manually setting a CPC score as following at simulation time: $T = 0$. This manual operation is hereafter called Variability-triggering event, creating a time-critical situation to cause variabilities of the functions. Here, it should be noted that detailed descriptions of the situation such as disappearing lane markings, construction zones, or merging motorway lanes, all of which are popular factors to build scenarios of the authority transfer problem, are avoided to cope with the envisioned world problem; the situation should be abstract rather than specific, and the CPCs in this context are adopted to simulate such abstract conditions.

Variability-triggering event: the score of CPC: “Available time” is set to 0 at simulation time $T = 0$, and the simulation process is triggered.

After a while of the Variability-triggering event, the potential couplings inherent to the prototype shown in Fig. 7 are further instantiated. This intervention into the simulation process is to represent a situation in which the ADS launches an alert and request the driver to take over the driving authority immediately, while the driver is still not aware of the emergency at all. This operation is hereafter referred to as Instantiation 1, for which three variations, i.e., Instantiation 1a, b and c are prepared; they are distinguished based on which level of the SA is supported by the warning system, and their details are shown in the following:

Instantiation 1a: the potential couplings from To ALERT DRIVERS TO PREPARE FOR EMERGENCY to To PAY ATTENTION TO TRAFFIC CONDITIONS and To PAY ATTENTION TO IN-CAR DISPLAY are instantiated from the instance shown in Fig. 7. This represents the launch of alerts, forcing the missing two functions related to the SA Lv. 1 to activate.

Instantiation 1b: the potential coupling from To ALERT DRIVERS TO PREPARE FOR EMERGENCY to To COMPREHEND CURRENT DRIVING SITUATION is instantiated. This represents the launch of alerts supporting the function related to the SA Lv. 2.

Instantiation 1c: the potential coupling from To ALERT DRIVERS TO PREPARE FOR EMERGENCY to To PLAN AND IDENTIFY NEXT ACTIONS is instantiated. This represents the launch of alerts supporting the function related to the SA Lv. 3.

In the end, additional instantiation is carried out after any one of the variations in Instantiation 1 to represent a situation in which the driver takes over the authority and resumes control. This intervention into the simulation process is hereafter referred to as Instantiation 2, and the detail is shown in the following:

Instantiation 2: To CONTINUE AUTONOMOUS DRIVING is deactivated, and To TAKE EVASIVE ACTIONS is activated; this represents the driver’s responses to the issue of the alert, i.e., disengagement of the ADS and following evasive actions.

By temporally combining the above instantiations, we can build a sequence of instantiations in three different ways, each of which corresponds to the difference in how and when the system supports the driver’s SA. This is shown in Table 5. Based upon these FRAM models corresponding to the three different scenarios, we simulate how the effect of the Variability-triggering event appears in each scenario and investigate into what the effective solutions for the authority transfer problem would be.

Here, it is actually possible to consider additional situations and elaborate the scenario something like: “what if there would be additional alarms after the initial TOR” or “what if the driver finds another threat to avoid (e.g., obstacles) as a result of road inspection, and the time-critical situation again during the evasive actions,” for example. The former scenario can easily be realized by manipulating the potential couplings coming from To ALERT DRIVERS TO PREPARE FOR EMERGENCY, and the latter scenario can be examined by setting the scores of CPCs: “Number of simultaneous goals” or “Available time” to 0, respectively after the Instantiation 2. Nevertheless, we try simple cases in this simulation because the scenario construction should not take into account too many conditions or change of parameters at once.

According to Hirose and Sawaragi (2020), one of the major limitations of their series of FRAM models is that the models essentially require complicated setting of parameters

Table 5 Simulation scenarios

Scenario no.	Flow of simulation scenario sequence
1	Initial state —> Variability-triggering event at $T = 0$ —> Instantiation 1a —> Instantiation 2
2	Initial state —> Variability-triggering event at $T = 0$ —> Instantiation 1b —> Instantiation 2
3	Initial state —> Variability-triggering event at $T = 0$ —> Instantiation 1c —> Instantiation 2

(e.g., functions, CPC weight, or simulation scenarios). In other words, too complicated parameter settings could easily lead to subjectivity or arbitrariness of simulations, and that is why the scenario should be constructed as simple as possible. Moreover, the focus of FRAM simulation is what kind of remarkable behavior can be observed based on the simple setting of parameters and their combinations, rather than collectively investigating complex cases; if there are still needs to consider more complicated scenarios, they should be extended based on the preceding—simpler simulations.

3.4 Simulation in each scenario and result

3.4.1 Simulation scenario 1

The first scenario is to investigate the effect of Variability-triggering event by the instantiations shown in Fig. 8. As stated in the previous subsection, the FRAM model of Fig. 5b was built by chronically merging Instantiation 1a and Instantiation 2 to original Instantiation 1. In the simulation, we varied the timings of Instantiation 1a and Instantiation 2 in various ways and investigate how the effects of the Variability-triggering event would change accordingly. That is, the timing of the alert, and the timing of the driver's invoking SA Lv. 1. All combinations of the timings to cause the Instantiation 1a and Instantiation 2 were simulated, and

Fig. 8 Instantiations in scenario 1

it was confirmed that the results could be classified into just three specific patterns shown in Fig. 9.

Figure 9a shows the transition pattern of the control modes in each function when Instantiation 1a and Instantiation 2 are activated at simulation time $T = 1$ and $T = 5$, respectively. Wherein, the horizontal axis represents the simulation time, and color gradations represent the control mode in each function. As this pattern shows, the successive

(a) Pattern 1: Instantiation 1a at $T = 1$ and Instantiation 2 at $T = 5$.

(b) Pattern 2: Instantiation 1a at $T = 6$ and Instantiation 2 at $T = 9$.

(c) Pattern 3: Instantiation 1a at $T = 6$ and Instantiation 2 at $T = 15$.

Fig. 9 Result of simulation scenario 1

instantiations result in the outcome that the driver can make an immediate recovery of all functions from the effect of the Variability-triggering event. This recovery trend can be commonly observed when both the Instantiation 1a and Instantiation 2 are activated earlier than the end of instantaneous peaks of control mode in TO PLAN AND IDENTIFY NEXT ACTIONS and TO TAKE EVASIVE ACTIONS, existing around the simulation time $T = 3$ –5. That is, the driver can recover from Variability, when both of the activations occur earlier than the simulation time $T = 5$ in this simulation.

Figure 9b shows the next emerging pattern of transition. The difference from the setting of Fig. 9a is the delay of instantiating Instantiation 1a and consequent Instantiation 2. That is, it was delayed to inform the driver of the authority transfer (i.e., take over request (TOR)), by which the driver can invoke his/her SA Lv. 1, and accordingly, the driver's taking evasive actions is also delayed. The result shows that it takes a longer time to recover from the Variability-triggering event than in the previous pattern. Moreover, the impact of the Variability-triggering event does never come to the stable within the time sequence of Fig. 9; an additional simulation confirms that the effect lasts until around the simulation time $T = 80$. This result shows that the delayed notification of the alert may cause a challenge to the driver to recover from the impact of the Variability-triggering event.

In addition to the above, another transition pattern emerges when the timing of Instantiation 2 is shifted to further later. The timing of Instantiation 1a is set the same with the previous case. In this pattern, the effect of the Variability-triggering event can be calmed down after the Instantiation 2 as shown in Fig. 9c. This trend can be observed even when the timing of Instantiation 2 was delayed to the timing when the control modes of the functions of TO PLAN AND IDENTIFY NEXT ACTIONS and TO TAKE EVASIVE ACTIONS are at the peaks (i.e., the most unstable), which exists around simulation time $T = 10$ –15. This suggests that the counter-measures of being notified to invoke the driver's SA Lv.1 (i.e., Instantiation 2) would be useful if it would be taken earlier than the driver's confusion comes to be maximum, even though the notification of the TOR (i.e., Instantiation 1a) was delayed. This is further discussed later.

3.4.2 Simulation scenario 2

The second scenario is to investigate the effect of the Variability-triggering event when the quality of support given by the system is changed from the previous case. The FRAM model used for this simulation is shown in Fig. 10. In this scenario, the role of the alarms is different from the previous scenario. It is supposed to provide some information telling the driver what is wrong and why, supporting the driver's SA Lv. 2. The driver needs to take some evasive actions based on it. The launch of the alarms is modeled as synchronized

Fig. 10 Instantiations in scenario 2

with the timings of Instantiation 1b and subsequent Instantiation 2. Wherein, Instantiation 2 is the same as assumed in the previous Scenario 1.

Figure 11a shows the case when Instantiation 1b and Instantiation 2 are caused at simulation time $T = 1$ and $T = 5$, respectively. The timing of the instantiations is precisely the same as the case shown in Fig. 9a. Also, the result is qualitatively the same, suggesting that the countermeasures referred to as the Instantiation 1b and Instantiation 2 can work effectively when they are taken in the early phase of the sequence. This trend lasts in the same way as the pattern 1 shown in Fig. 9a. However, the emerging behaviors become quite different from the previous scenario shown in Fig. 9 when the alarms for the SA Lv. 2 come in delay.

Figure 11b illustrates another transition pattern emerging in this scenario, wherein the timings of Instantiation 1b and Instantiation 2 are set the same with the case of Fig. 9c. However, the result is qualitatively different from the previous scenario drastically. The effect of the Variability-triggering event does never come to rest forever, even after the

countermeasures are taken. This trend is similarly observed throughout the other combinations of timings to instantiate. All the results obtained in this scenario can eventually be classified into only two patterns shown in Fig. 11. According to the result, there exist chances to recover from the effect of the Variability-triggering event only when the instantiations were done in the early phase described in Fig. 9a. Besides, any chances of recovery would be lost if they were done in delay. This suggests that the direct support of the SA Lv. 2 by the automation might be less effective than the support of the SA Lv. 1 in too time-critical situations.

3.4.3 Simulation scenario 3

Figure 12 shows the instantiations in this scenario. The role of alarms in this scenario is to provide some information instructing the driver about actual evasive actions, supporting the SA Lv. 3 and their DM. The drivers are supposed to follow them. The launch of the alarms is represented as the

Fig. 11 Result of simulation scenario 2

Fig. 12 Instantiations in scenario 3

Instantiation 1c shown in Fig. 12a, and the evasive action is also instantiated as shown in the figure.

Figure 12 shows the instantiations in this scenario. The role of alarms in this scenario is to provide some information instructing the driver about actual evasive actions, supporting the SA Lv. 3 and consequent DM of evasive actions. The launch of the alarms is represented as the Instantiation 1c shown in Fig. 12a, and the subsequent evasive action is modeled by Instantiation 2 as shown in Fig. 12b. All combinations of timing to cause the Instantiation 1c and Instantiation 2 are simulated just like the other two scenarios. It was confirmed that the trend of the emerging transition patterns in this scenario are qualitatively the same as ones observed in the previous scenario of Fig. 11. That is, there would exist chances to recover from the effect of the Variability-triggering event only when the alarms come earlier, but the chances of the recovery were lost entirely otherwise. This also suggests that the direct support of the SA Lv. 3 by the automation might be less effective in too time-critical situations as well.

3.5 Relationships between simulation results and variation in Instantiation 1

According to the previous subsection, the simulation result in each scenario can be classified into two categories: the one is the simulation scenario 1 in which the countermeasure succeeds in the recovery from the effect of Variability-triggering event; the other is the simulation scenario 2 and 3 in which any chances of the recovery would be lost if the countermeasures are taken in delay. Now that the crucial difference among the simulation scenarios can be seen in the variation of the Instantiation 1, the variation seems to be related to the difference of the result. An additional simulation was conducted to verify this, and the effect of Instantiation 1a, 1b, and 1c is investigated here.

This simulation is, more specifically, to examine characteristics of the instances shown in Figs. 8a, 10a, and 12a in response to the effect of the Variability-triggering event. In this simulation, each of Instantiation 1a, 1b, or 1c was carried out first of all at the simulation time $T = 0$, and then the Variability-triggering event was also caused at $T = 0$. There were no interventions into the simulation process (e.g., Instantiation 2) after that, and the transition of the control mode was simulated until $T = 60$.

Figure 13a and b shows the result when the Instantiation 1a and 1b are carried out at $T = 0$, respectively; it should be noted that the result in the case of the Instantiation 1c is qualitatively the same as shown in Fig. 13b, so we omit this case. Now, we can confirm in Fig. 13a that the effect of the Variability-triggering event is calmed down in the case of the resultspsonlyInstantiation1(b), but keeps on remaining in the case of the resultspsonlyInstantiation1(a); further simulation confirmed that the effect of Variability-triggering event lasts until around the simulation time $T = 700$ in the case of the Instantiation 1b.

This result can be explained from the perspective of static and dynamic stability of the FRAM model, which is usually discussed in aerodynamics of flight (FAA 2016). The static stability in this context refers to a potential characteristic of a system against disturbances, depending on its inherent component such as structure or design; the static stability is positive when a system returns to the original state of equilibrium after being disturbed, neutral when a system remains in a new condition after its equilibrium has been disturbed, and negative when a system continues away from the original state of equilibrium after being disturbed. On the other hand, the dynamic stability refers to a time-dependent response of a system based on its static stability, disturbances, and additional input signals; the dynamic stability is positive when a displaced system decreases its amplitude of the displacement and returns toward an equilibrium state

Fig. 13 Responses of the instances shown in Figs. 8a and 10a against the effect of the Variability-triggering event

over time, neutral when the displacement neither decreases nor increases in amplitude, and negative when the amplitude of the displacement increases and becomes more divergent. Therefore, the result found that the static stability of the instance shown in Fig. 8a is positive, and that of the instance shown in Fig. 10a is substantially neutral since its recovery time is more than ten times longer than the case of the Instantiation 1a.

The simulation result further suggests the role of each instantiation: the variation of Instantiation 1s play a role in transforming the static stability of the FRAM model against the effect of the Variability-triggering event; Instantiation 2 has effect on the dynamic stability of the FRAM model—or how fast the effect Variability-triggering event is calmed down. The dynamic stability generally exists only in the condition of the positive static stability, and the static stability was changed from neutral to positive only in the case of Instantiation 1a. That is why the composite effect of Instantiation 1a and 2 succeeded in suppressing the effect of Variability-triggering event, and the effect of Instantiation 1b or 1c and 2 failed in the late phase of the simulation sequence. Therefore, from the perspective of the functional abstraction level, how to establish the positive static stability of the FRAM model is the highest priority, and how and when to respond to TOR should be of secondary importance.

3.6 Brief summary of simulations

We confirmed that all FRAM models for the three simulation scenarios can work well and are expected to contribute to verifying how an event of the authority transfer affects the performance of the driving and what the restrictions for those. Furthermore, a countermeasure to recover from performative variabilities caused by the authority transfer was also validated. This was consistent with our intuition because we have been simulating the authority transfer in a time-critical situation, and the result suggests that the faster is better. Also, the result gained from the FRAM model of Fig. 8 verified that the system's notification at the level of a driver's SA Lv. 1 turned out to be effective even after the instruction of TOR to a driver is delayed, and the recovery mechanism is related to the static stability of the FRAM model at the functional abstraction level.

This is crucial because the reaction time of human drivers is generally limited, and it is less likely for the drivers to react immediately to the abrupt instruction. Notably, the reaction time could be even longer if they were occupied with some other secondary tasks.

This leads us to conclude that the driver's SA should be left in the control loop, and the system's support for the driver's SA should be designed so that the drivers could be aware of the situation from Lv. 1 by themselves. On the other hand, if the system's support for the driver's SA starts from

Lv. 2 and Lv. 3, the confusion of the driver might be more, and they could not be adequate to let them recover from the chaos. The possible reason for this is that the information related to the SA Lv. 2 and 3 is generally so ad hoc that it does not support their contextual awareness and strategic actions. This results in the driver's opportunistic reactions, and any chances of the recovery are lost in the end. Therefore, the establishment of SA Lv. 1 could play a significant role in conditional driving automation.

4 Discussions: feasibility of conditional driving automation

The simulation result suggests that the drivers should pay attentions to the driving conditions even if they are allowed to be free from the dynamic driving tasks (DDTs) by the conditional driving automation or LoDA 3. However, the history of automation has been showing that it is intrinsically very burdensome for the human beings to keep on monitoring the systems and maintain their SA. This human characteristic has been proved by the professionals of the automation such as airline pilots or operators of nuclear power plants, and it is not difficult to imagine that the situation could be even worse in the case of non-professionals of the automation (i.e., usual human drivers). Therefore, it is now questioning whether the conditional driving automation is feasible in the real field of practice. Edwards et al. (2016), Dehais et al. (2012), and Inagaki and Sheridan (2019) discussed the important points related to this problem.

Edwards et al. (2016) have experimented with investigating the relationship between workload modified by the automation support, SA, and responding time, i.e., performance of air traffic controllers. They found through their experiment that the too much support of the automation could result in a reduction of engagement, leading to reduced SA and performance. They eventually concluded that the automation should not be designed just to reduce the workload; the automation should be designed to support the human operators to maintain their contextual SA.

In terms of the contextual SA, Dehais et al. (2012) have also conducted an experiment to investigate how conflicts of intentions between human operators and automated systems affect the human operators' performance. The result has confirmed that the conflicts occurred during the experimental operation of an unmanned ground vehicle (UGV) led to perseveration behavior, higher heart rate, excessive-limited attentional focus of a major part—69.2% of the participants, having resulted in the failure of their given task. They pointed out based on this result that automation overriding human operators' actions such as protection systems could cause this kind of conflicts and perseveration behaviors. Moreover, such operational supports or interventions

by the automation alone are meaningless as long as the human operators are out of the loop and does not understand their behavior. The priority of automation design should, therefore, be put more on how to “cure” persevering human operators when the operators face a conflict, for which one possible solution is to identify and control the conflicts so that the human operators reconstruct their persevering SA.

Inagaki and Sheridan (2019) have pointed out the possibility that the drivers could fail to respond to a TOR during the autonomous driving at LoDA 3, for which additional operational supports such as steering or applying brake by the automation might be necessary. They proved this with a mathematical evaluation, suggesting that the smooth authority transfer cannot be achieved without some highly automated operational supports. On the other hand, they also discussed that this evaluation is no longer consistent with the framework of the LoDA 3, which cannot be classified into the single category; it should be noted that the evaluation also conflicts with the suggestions provided by Dehais et al. (2012). In the end, they concluded that the LoDA 3 is not the level where the R&D of the autonomous driving should aim and instead proposed two solutions: the one is to revise the definition of the conditional driving automation issued by SAE (2016); the other is to design HMI enhancing the human–machine collaboration so that the driver can smoothly take over the authority.

Our conclusion is consistent with theirs: the drivers must still maintain good SA during the autonomous driving while the system should also issue the TOR with its highly automated operational supports, if necessary. This suggests that human-machine collaboration still plays a significant role there, and the nature of the authority transfer problem is ambiguous rather than explicit. Both humans and machines are always responsible for driving, and it is irrational to consider the role of them individually. The focus should, therefore, be put on how to realize the productive human–machine collaboration throughout the driving sequence.

As an example, Cropley (2019) has reported that speed limiters and driver monitors will be mandatory in EU from 2022 and pointed its negative aspects from the perspective of human cognitions, all of which are consistent with the discussions so far. No matter how the importance of productive human–machine collaboration has been stressed for decades, the reality is that people still want to entrust everything to technologies in this way. This kind of approach actually sounds reliable and socially acceptable at a glance since human beings are still regarded as something like “sources of errors.” However, our simulation result suggests that this is no longer true. The productive human–machine collaboration must not be neglected, no matter how sophisticated automation supports are introduced.

One possible solution for the productive human–machine collaboration is to develop an HMI or some other

mechanisms with which the drivers can keep their attention to the driving. The simulation result has shown that the SA should be established without missing any levels. Moreover, the establishment of the SA Lv. 1 must be supported by the system with the highest priority, and the more automated cognitive support for the SA Lv. 2 or 3 should follow it even if necessary; otherwise, the performance of the drivers could be reduced rather than enhanced. The support of the SA Lv. 1 can be achieved without any complicated, high technological equipment since the drivers only have to notice trivial elements related to the current situation. Specifically, they just need to know or feel “everything is going well” or “something is wrong” during the autonomous driving and the further SA should be developed by themselves.

In this respect, Vanderhaegen (2016) and Vanderhaegen (2017) has addressed the related issue from the perspective of cognitive dissonance (Festinger 1957). The cognitive dissonance is an incoherence among cognitions of human beings. The dissonance occurs when something seems to be wrong or different, and we feel gaps or conflicts among our individual or collective knowledge; it may generate our discomfort or a situation overload, and the discomfort may recursively result in further dissonance, whose process enables us to discover new knowledge or cognitions. In this context, Vanderhaegen (2016) proposed an original rule-based tool to model human and technical behaviors, detect possible conflicts among them, and assist the discovery and control of the dissonance; it should be noted that this approach can be associated with the identification and control of the conflicts suggested by Dehais et al. (2012). In addition, his succeeding study (Vanderhaegen 2017) suggested that such unstable conditions, i.e., dissonance could contribute to adaptive behaviors of human beings or the resilience of systems (Holling 1973; Hollnagel et al. 2006; Hollnagel 2017).

An aspect of this approach can be confirmed in our result as well. The result shown in Fig. 9c suggests that the instantaneous peak of the control mode, i.e., unstable condition in two functions: TO PLAN AND IDENTIFY NEXT ACTIONS and TO TAKE EVASIVE ACTIONS could contribute to the immediate recovery from the effect of variabilities. This result is therefore consistent with the concept of dissonance in terms of using unstable for stable/resilience. It is essential to take into account these aspects for the productive human–machine collaboration, and the future work will be conducted to elucidate its nature.

5 Conclusion

The technology of autonomous driving has rapidly been developed, aiming at conditional driving automation or LoDA 3. One of the most critical challenges for the achievement is how to take over the driving authority in

an emergency. Many researchers from both academic and industrial fields have been working on this issue, but any fundamental solutions have not been found out yet.

To investigate the nature of the authority transfer in time-critical situations, a simulation based on FRAM was carried out. The result has shown that the driver's SA should be left in the control loop even during the autonomous driving. Specifically, the system's support for the driver's SA should be designed so that the drivers could be aware of the situation from Lv. 1 by themselves, otherwise any chances of the recovery in the authority transfer sequence would be lost. This also corresponds to transforming the static stability of the FRAM model from neutral to positive at the functional abstraction level, and maintaining the positive static stability should be the highest priority. Therefore, the positive static stability plays a significant role in the authority transfer problem, and it is useless to try any measures without considering the static stability of the functional model; the transformation of the static stability is triggered by the establishment of SA Lv. 1, according to this simulation result.

While the simulation result suggests that the driver's SA should be established without missing any levels, the history of automation has been showing that it is intrinsically very burdensome for the human beings to keep on monitoring the systems and maintain their good SA. The focus should, therefore, be put on how to realize the productive human–machine collaboration throughout the driving sequence, and it is irrational to consider the role of them individually; the human must be more or less engaged in DDTs anytime as long as the full—100% automation was not invented.

In the end, the safety of the authority transfer in time-critical situation has been simulated at the functional abstraction level, and correspondences between the abstracted, functional world and the real, physical world has been investigated throughout this paper. Human–machine interactions in the real field of practice are generally so complex that it is quite difficult to properly envision actual situations and practical requirements in the R&D process. Moreover, there are some limitations to conduct the R&D of new technologies with physical or real-oriented approaches only, which is the nature of the envisioned world problem. The approach from the multi-abstraction levels which has been shown in this paper is a solution to address the problem; its importance must increase in near future as the rapid evolution of socio-technical systems.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated

otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Brauer W, Reisig W (2009) Carl adam petri and “Petri Nets”. *Fundam Concepts Comput Sci* 3(5):129–139
- Bytheway CW (2007) FAST creativity and innovation: rapidly improving processes. J Ross Publishing, Product Development and Solving Complex Problems
- Cropley S (2019) Speed limiters may create more dangers than they prevent: EU's decision to make speed limiters mandatory on all new cars from 2022 plays to the view that we're overly reliant on car tech. <https://www.autocar.co.uk/opinion/industry/speed-limiters-may-create-more-dangers-they-prevent>. Accessed 4 Aug 2020
- Dehais F, Causse M, Vachon F, Tremblay S (2012) Cognitive conflict in human-automation interactions: a psychophysiological study. *Appl Ergon* 43(3):588–595
- Edwards T, Homola J, Mercer J, Claudatos L (2016) Multifactor interactions and the air traffic controller: the interaction of situation awareness and workload in association with automation. *IFAC Pap Online* 49(19):597–602
- Endsley MR (1995) Toward a theory of situation awareness in dynamic systems. *Hum Factors* 37(1):32–64
- Eriksson A, Stanton NA (2017) Takeover time in highly automated vehicles: noncritical transitions to and from manual control. *Hum Factors* 59(4):689–705
- FAA (2016) Pilot's handbook of aeronautical knowledge. *Handbooks & Manuals FAA-H-8083-25B*, Federal Aviation Administration
- Festinger L (1957) A theory of cognitive dissonance, vol 2. Stanford University Press, Palo Alto
- Hirose T, Sawaragi T (2019) Development of FRAM model based on structure of complex adaptive systems to visualize safety of socio-technical systems. *IFAC Pap Online* 52(19):13–18
- Hirose T, Sawaragi T (2020) Extended FRAM model based on cellular automaton to clarify complexity of socio-technical systems and improve their safety. *Saf Sci* 123
- Holling CS (1973) Resilience and stability of ecological systems. *Annu Rev Ecol Syst* 4(1):1–23
- Hollnagel E (2003) Context, cognition and control. *Coop Process Manag* 27–52
- Hollnagel E (1998) Cognitive reliability and error analysis method (CREAM). Elsevier, Amsterdam
- Hollnagel E (2004) Barriers and accident prevention. Ashgate Publishing Ltd, Farnham
- Hollnagel E (2012) FRAM: the functional resonance analysis method: modelling complex socio-technical systems. Ashgate Publishing Ltd, Farnham
- Hollnagel E (2017) Safety-II in practice: developing the resilience potentials. Taylor & Francis, New York
- Hollnagel E, Woods DD, Leveson N (2006) Resilience engineering: concepts and precepts. CRC Press, Boca Raton
- Inagaki T (2003) Adaptive automation: sharing and trading of control. *Handb Cogn Task Des* 8:147–169
- Inagaki T, Sheridan TB (2019) A critique of the SAE conditional driving automation definition, and analyses of options for improvement. *Cogn Technol Work* 21(4):569–578

- Konstandinidou M, Nivolianitou Z, Kiranoudis C, Markatos N (2006) A fuzzy modeling application of CREAM methodology for human reliability analysis. *Reliab Eng Syst Saf* 91(6):706–716
- Kuipers B (1986) Qualitative simulation. *Artif Intell* 29(3):289–338
- Kuipers B (1994) Qualitative reasoning: modeling and simulation with incomplete knowledge. MIT press, Cambridge
- Kuipers B (2001) Qualitative simulation. *Encycl Phys Sci Technol* 3:287–300
- Lee WS, Grosh DL, Tillman FA, Lie CH (1985) Fault tree analysis, methods and application – a review. *IEEE Trans Reliab* 34(3):194–203
- Lind M (2011) An introduction to multilevel flow modeling. *Nucl Saf Simul* 2(1):22–32
- Neisser U (1976) Cognition and reality. W. H. Freeman, New York
- Petri CA (1966) Communication with automata. Technical Report RADC-TR-65-377, Griffiss air force base, New York
- Rasmussen J (1986) Information processing and human-machine interaction: an approach to cognitive engineering. Elsevier Science Publishing Company, Inc.
- Rasmussen J, Pejtersen AM, Goodstein LP (1994) Cognitive Systems Engineering. Wiley, Hoboken
- Ross DT (1977) Structured analysis (sa): a language for communicating ideas. *IEEE Trans Softw Eng* 1:16–34
- Roth EM, Mumaw RJ (1995) Using cognitive task analysis to define human interface requirements for first-of-a-kind systems. *Proc Hum Factors Ergon Soc Annu Meet* 39(9):520–524
- Saaty TL (1990) How to make a decision: the analytic hierarchy process. *Eur J Oper Res* 48(1):9–26
- SAE (2016) Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. *SAE Int (J3016)*
- Smith K, Hancock PA (1995) Situation awareness is adaptive, externally directed consciousness. *Hum Factors* 37(1):137–148
- Stamatelatos M, Vesely W, Dugan J, Fragola J, Minarick J, Railsback J (2002) Fault tree handbook with aerospace applications. Handbook, NASA
- Swain AD, Guttman HE (1983) Handbook of human-reliability analysis with emphasis on nuclear power plant applications. Final report. Tech. rep, Sandia National Labs
- Ung ST (2015) A weighted CREAM model for maritime human reliability analysis. *Saf Sci* 72:144–152
- Vanderhaegen F (2016) A rule-based support system for dissonance discovery and control applied to car driving. *Expert Syst Appl* 65:361–371
- Vanderhaegen F (2017) Towards increased systems resilience: new challenges based on dissonance control for human reliability in Cyber-Physical&Human Systems. *Annu Rev Control* 44:316–322
- Vesely WE, Goldberg FF, Roberts NH, Haasl DF (1981) Fault tree handbook. Handbook NUREG-0492, U.S. Nuclear Regulatory Commission
- Vicente KJ (1999) Cognitive work analysis: toward safe, productive, and healthy computer-based work. CRC Press, Boca Raton
- Woods DD, Christoffersen K (2002) Balancing practice-centered research and design. In: Cognitive Systems Engineering in Military Aviation Domains Wright-Patterson AFB: Human Systems Information Analysis Center, pp 121–136
- Woods DD, Dekker S (2000) Anticipating the effects of technological change: a new era of dynamics for human factors. *Theor Issues Ergon Sci* 1(3):272–282
- Yang Z, Bonsall S, Wall A, Wang J, Usman M (2013) A modified CREAM to human reliability quantification in marine engineering. *Ocean Eng* 58:293–303

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.