



# A critique of the SAE conditional driving automation definition, and analyses of options for improvement

Toshiyuki Inagaki<sup>1</sup> · Thomas B. Sheridan<sup>2</sup>

Received: 10 August 2017 / Accepted: 11 February 2018  
© The Author(s) 2018. This article is an open access publication

## Abstract

The Society of Automotive Engineers defines five *levels of driving automation* (LoDA) (plus a “no-automation” level 0). Among them, the third level, called “conditional driving automation,” here denoted LoDA 3, performs the complete *dynamic driving task* (DDT) within a limited operational domain. Although the driver is free from any driving task while the automation is engaged, she is expected to be receptive to an automation-issued *request to intervene* (RTI) and is also expected to perform DDT fallback in a timely manner. This paper gives a method to derive an optimal design for RTI and proves that LoDA 3 coupled with the optimal RTI should never be simply called “conditional driving automation.” This means that the definition of LoDA 3 is not complete and that at least one important level is missing in the list for LoDAs. This paper provides two ways to resolve the problem.

**Keywords** Automated driving · Levels of driving automation · Request to intervene · Levels of automation · Trading of authority · Shared control

## 1 Introduction

The term “automated driving” has been attracting keen interest worldwide. However, the term can have many different meanings. Actually, several varieties can be distinguished for automated driving, depending on the scheme of function allocation between the driver and the automation. If the driver forms incorrect mental models of the function allocation, various human factor-related problems will arise.

Considerable effort has been made to distinguish categories of automated driving by institutions, such as the German Federal Highway Institute (BASt 2013; Gasser and Westhoff 2012), the National Highway Traffic Safety Administration (NHTSA 2013), and the Society for Automotive Engineers (SAE 2016). Among them, the SAE J3016 standard’s

definition (SAE 2016) for *levels of driving automation* (LoDA) seems to be gaining popularity as a common language to be used worldwide. SAE J3016 distinguishes five LoDAs (Table 1). *Driver Assistance* has been used for many years. *Partial Driving Automation* is expected to be put into practical use within a few years. *High Driving Automation* or *Full Driving Automation* may need some more years before legal questions are solved appropriately in related countries. Efforts are focused mainly now on *Conditional Driving Automation* or LoDA 3.

SAE J3016 defines LoDA 3 as follows: The automation performs the complete *dynamic driving task* (DDT), including lateral and longitudinal control as well as object and event detection and response (SAE 2016). While the automation is engaged, the driver is free from any driving task and may be involved in non-driving tasks. However, when the automation detects that operational design domain limits are about to be exceeded or that there is some DDT performance-relevant system failure, the automation issues a *request to intervene* (RTI), with the expectation that the driver notices the RTI and intervenes in a timely manner to rectify the situation. In other words, the driver is assumed to be receptive to an automation-issued RTI and be ready to perform DDT fallback (SAE 2016). The SAE-defined level 3 driving automation disengages an appropriate time

✉ Toshiyuki Inagaki  
inagaki.toshiyuki.gb@un.tsukuba.ac.jp  
Thomas B. Sheridan  
Sheridan@mit.edu

<sup>1</sup> Laboratory for Cognitive Systems Science, University of Tsukuba, Tsukuba 305-8577, Japan

<sup>2</sup> Departments of Mechanical Engineering and Aeronautics/Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, USA

**Table 1** SAE levels of driving automation (LoDA)

Level of driving automation (LoDA)	Definition
LoDA 1: driver assistance	Automation performs either longitudinal or lateral vehicle motion control (on a sustained basis), but not complete <i>object and event detection and response</i> (OEDR)
LoDA 2: partial driving automation	Automation performs both longitudinal and lateral vehicle motion control (on a sustained basis), but not complete OEDR
LoDA 3: conditional driving automation	Automation performs the complete <i>dynamic driving task</i> (DDT), but not DDT fallback, within a limited <i>operational design domain</i> (ODD)
LoDA 4: high driving automation	Automation performs the complete DDT and DDT fallback within a limited ODD
LoDA 5: full driving automation	Automation performs the complete DDT and DDT fallback without ODD limitation

After (SAE 2016)

In (SAE 2016), *dynamic driving task* (DDT) is defined as a collection of the following five subtasks: (1) lateral vehicle motion control, (2) longitudinal vehicle motion control, (3) monitoring the driving environment via object and event detection, recognition, classification, and response preparation, (4) object and event response execution, (5) maneuver planning, and (6) enhancing conspicuity via lighting and gesturing, etc., in which (3) and (4) are referred to as *object and event detection and response* (OEDR)

after issuing an RTI, which may impose time pressure on the driver. Many studies have been made to investigate takeover-related issues, such as automation-to-driver transitions of control (Lu et al. 2016), time allowance for the driver to take over (Gold et al. 2013; Zeeb et al. 2015), driver's takeover performance and its quality (Merat et al. 2014; Zeeb et al. 2016; Happee et al. 2017), and RTI interface design to assist the driver's maneuver (Petermeijer et al. 2017).

However, there seem to be few studies discussing what kind of action the automation should take when the driver does not respond properly to an RTI. Is it realistic to assume that the driver is always receptive to an automation-issued RTI? Is it sensible to expect that the driver is DDT fallback-ready no matter what are the circumstances when automation issues an RTI? The driver might fail to notice that an RTI was issued, or might fail to grasp and interpret the situation properly. The driver might be unable to determine what to do in the situation, or might be unable to perform what she decided to do. This suggests that an RTI must be designed with care by taking into account possibilities that the driver can fail to respond to a given RTI within a time limit. In such investigations, the following questions should be asked: "What needs to be communicated to the driver in the RTI?"; "What may the driver do when she feels difficulty in taking over control from the automation?"; and "What should the automation do when the driver does not respond effectively to an RTI?"

This paper proposes a systematic method to generate a list of design alternatives of RTI for SAE LoDA 3. It derives an "optimal" RTI that yields a maximal value for a utility function defined as the weighted sum of the benefit derived from successful taking over and the cost arising out of the state in which the vehicle is controlled neither by the automation nor by the driver. We show that LoDA 3 coupled with the optimal RTI should never simply be called "conditional driving automation," i.e., as a single category. This means

that Table 1 is incomplete as the list for LoDAs, and that an important level is missing there. This paper gives two possible ways to resolve the problem: One is to include one of the levels discussed in (SAE 2014), the previous version of SAE J3016 published in 2016, and the other is to revise the definition of LoDA 3 in (SAE 2016) appropriately. Otherwise, LoDA 3 may not be a reasonable objective to be translated into reality.

## 2 Designing messages for a request to intervene (RTI)

This section proposes that the generic "*levels of automation*" (LoA) concept that has been in the literature for many years is useful in obtaining various design alternatives for RTI messages. The LoA concept was introduced originally by Sheridan and Verplank (1978) to suggest a wide variety of automation modes that replace functions previously carried out by the human. It has been applied to various human-machine systems in the real world, such as teleoperation, aviation, surface transport, nuclear and process plants, manufacturing, mobile robots (Inagaki 2005; Sheridan 1992; Endsley and Kaber 1999; Inagaki and Inoue 1993; Kaber and Endsley 2007; Lindstrom and Winroch 2010; Inagaki et al. 2007; Moray et al. 2000; Pacaux et al. 2011; Parasuraman et al. 2000; Jou et al. 2009).

### 2.1 Levels of automation

Table 2 shows LoA in an 11-point scale, ranging from LoA 1 through LoA 10, with LoA 6.5 (Inagaki et al. 1998) between LoA 6 and LoA 7 in the original 10-point scale defined in Sheridan (1992, 1999).

We believe that NHTSA (2013, 2016) and SAE (2014), which is an earlier version of (SAE 2016), confuse the more

**Table 2** Levels of automation (LoA)

LoA 1	The computer offers no assistance: the human must make all decisions and actions
LoA 2	The computer offers a complete set of decision/action alternatives, or
LoA 3	narrows the selection down to a few, or
LoA 4	suggests one alternative, and
LoA 5	executes that suggestion if the human approves, or
LoA 6	allows the human a restricted time to veto before automatic execution, or
LoA 6.5	executes automatically after telling the human what it is going to do, or
LoA 7	executes automatically, then necessarily informs humans, or
LoA 8	informs the human after execution only if asked, or
LoA 9	informs the human after execution only if it, the computer, decides to
LoA 10	The computer decides everything and acts autonomously, ignoring the human

After Sheridan (1992, 1999), Inagaki et al. (1998)

general “levels of automation” and the “levels of driving automation” (explicit to driving) and use the phrases interchangeably. Some of the above-mentioned papers discussing the takeover-related issues inherit such a wording. However, as Tables 1 and 2 tell clearly, the two concepts are quite different and should be distinguished appropriately. In this paper, we analyze the differences in the implications of the scales, particularly at the LoA levels that seem to relate to LoDA 3. Our analysis shows that the LoA concept can deepen our understanding of LoDA, and the LoDA definition, on the other hand, can give an opportunity to re-realize the power of LoA.

## 2.2 Who is in authority?

In the concept of *human-centered automation*, it is usually assumed that the human is maintained as the final authority over the automation (Billings 1997; Sheridan 2002; Woods 1989). LoAs positioned at 5 or below are compatible with the human-centered automation principle, because the human is always in authority. Aviation is a typical domain in which human-centered automation is accepted (Billing 1997); see Example 1.

**Example 1** The Traffic alert and Collision Avoidance System (TCAS) is an automated system to reduce risks of mid-air collision by issuing *resolution advisories* (RA) with an aural message to tell the pilot whether to climb or descend. Although the pilot is expected to respect an RA, she may disobey it if there is a definite reason. The LoA of TCAS in this case is positioned at 4. When TCAS issues an RA, the pilot must disconnect the autopilot as well as the flight directors and then adjust the pitch attitude of the aircraft so that the vertical speed suggested by the RA may be attained. It is known that such a rarely used flying technique imposes a heavy burden on pilots (Botagargues 2009). For a TCAS that performs the maneuver recommended by the RA where the

pilot expresses agreement with it, the LoA of such a TCAS can be said to correspond to LoA 5.

When the LoA is positioned at 6 or higher, the situation becomes different. In case of LoA 6, the automation allows the human a limited time to veto its proposed action. Here the automation executes the action if the allotted time passes while the human is dithering whether or not to veto. In this case, the human is not treated as the final authority over the automation. However, that does not mean that higher levels of automation are unnecessary. Even in aviation, as a matter of fact, such systems are in use to cope with situations in which immediate and precise control maneuvers are indispensable to assure safety; see Examples 2–4, again based on experience in aviation technology.

**Example 2** Some aircraft are equipped with an automated system that can cope with cabin decompression in a highly automatic manner (LoA 6). When the aircraft’s monitoring system detects unsafe cabin pressure, it gives the crew a warning and begins a countdown. If the crew neither cancels the warning nor applies any positive control action to the aircraft, the automation performs a side-step maneuver followed by an automatic rapid descent (Kaminski-Morrow 2009).

**Example 3** A new type of TCAS provides necessary information and directives to the autopilot so that it can control the aircraft automatically and precisely as the RA demands, so as to avoid a mid-air collision (Botagargues 2009). The new TCAS does not need any human intervention, which means that LoA of the new TCAS is positioned at 6.5 or higher.

**Example 4** The *Thrust Asymmetry Compensation* (TAC) for a twin-engine aircraft is an automated system with LoA 7. During a takeoff roll, TAC monitors engine data continually to determine the thrust level of each engine. If TAC detects

a sufficient difference between the two engines, it automatically controls the rudder to reduce skidding-off-runway accidents. The “*rudder movement is back-driven through the rudder pedals and the rudder trim indicator to provide rudder control awareness to the pilot*” (Boeing 2005).

### 2.3 Designing a request to intervene in driving

SAE J3016 (2016) defines that LoDA 3 performs the complete DDT, but does not carry out DDT fallback. The automation issues a timely RTI to the DDT fallback-ready user when it determines that operational design domain limits are about to be exceeded or there is a DDT performance-relevant system failure. SAE J3016 (2016) does not specify RTI messages explicitly. However, it assumes that the automation disengages an appropriate time after issuing an RTI. A first candidate for an RTI message might thus be, “Intervene and resume driving within  $T$  seconds time,” which we here call Baseline.

In designing RTI messages, it would be realistic to assume that the driver either might not be receptive to an RTI or be ready to perform the fallback maneuver. This section illustrates how design alternatives for an RTI can be obtained systematically by consulting Table 2.

LoAs that are positioned at either 3 or lower may be excluded from our discussion because they are not appropriate for asking a single action to take (e.g., “resume driving”). For instance, LoA 3 is useful when the computer shows the human a few candidate actions so that she can pick up one among them. However, in our case of an RTI, there are no candidate actions other than to “resume driving.” LoA 4 is not useful, either, in designing an RTI message: Although the computer recommends the driver to “resume driving,” she has the right to disregard the recommendation to take a completely different action that she likes.

LoA 5 when applied to RTI automation implies an RTI message that asks the driver to intervene and resume driving within  $T$  seconds time. It also tells the driver that the automation disengages when it confirms that the driver resumes driving. Namely, the automation suggests the driver, “You intervene and resume driving. Then I disengage myself.” This means that the automation does not disengage when the driver does not respond to an RTI. The automation continues vehicle operation within its power to lead the vehicle

possibly to a minimal risk condition (e.g., stopping on the road or the shoulder).

LoA 6 implies another design alternative of an RTI message that asks the driver to intervene and resume driving within  $T$  seconds time. It also tells the driver that she may veto when she is not willing to resume driving due to her lack of confidence in taking care of the situation. If the RTI is vetoed, the automation continues vehicle operation to lead the vehicle to a minimal risk condition. If no response is given to the RTI, the automation disengages when  $T$  seconds have passed. In that case, neither the driver nor the automation controls the vehicle.

LoA 6.5 suggests a design alternative of an RTI message that tells the driver to intervene and resume driving at once and that the automation is going to disengage shortly. If the driver fails to respond to the RTI in a timely manner, the vehicle is put into a condition in which nobody controls it. Even if the driver resumes driving, the driver’s vehicle operation may not be of adequate quality because the takeover time is extremely short.

LoAs positioned at either 7 or higher may not be suitable for an RTI message, because they do not communicate with the driver at all before the automation disengages.

Table 3 gives design alternatives for an RTI message that may be used in LoDA 3 of the SAE system.

## 3 Evaluating design of a request to intervene message

Table 4 summarizes the driver’s response to various RTI messages and consequences that may occur.

### 3.1 Expected utility for an RTI

Let  $P(\text{RD}|\text{Baseline})$  denote the probability that the driver resumes driving (RD) after she is requested to intervene within  $T$  seconds time, and  $P(\text{NR}|\text{Baseline})$  be the probability of no response (NR) to the request and thus the automation disengages when  $T$  seconds have passed. The driver’s response to the RTI is either RD or NR, namely  $P(\text{RD}|\text{Baseline}) + P(\text{NR}|\text{Baseline}) = 1$ . Let  $a$  denote the benefit of successful fallback by the driver, and  $c$  the cost arising out of the state in which the vehicle is controlled

**Table 3** Design alternatives of an RTI message for LoDA 3

LoA of RTI	Message in RTI
LoA 5	“Intervene and resume driving within $T$ seconds time. The automation disengages when it confirms that you have started vehicle operation.”
LoA 6	“Intervene and resume driving within $T$ seconds time. Veto if you do not like to do so.”
LoA 6.5	“Intervene and resume driving at once. The automation disengages soon.”
Baseline	“Intervene and resume driving within $T$ seconds time.”

**Table 4** Human response to an automation-issued RTI and consequences that may occur

LoA of RTI	If the human accepted the RTI	If the RTI was given no response or if RTI was vetoed in the case of LoA 6
LoA 5	The human resumes vehicle operation	The automation may not disengage and must continue vehicle operation within its power to lead the vehicle possibly to a minimal risk condition
LoA 6	The human resumes vehicle operation	If the RTI was vetoed, the automation continues vehicle operation within its power to lead the vehicle possibly to a minimal risk condition If no response was given to the RTI, the automation disengages when $T$ seconds have passed. Nobody controls the vehicle after that
LoA 6.5	The human resumes vehicle operation that may not be good in quality due to short takeover time	Nobody controls the vehicle because the automation disengages immediately after issuing the RTI
Baseline	The human resumes vehicle operation	The automation disengages when $T$ seconds have passed. Nobody controls the vehicle after that

neither by the automation nor the driver. The expected utility  $U(\text{Baseline})$  for an RTI of Baseline type is given by:

$$U(\text{Baseline}) = aP(\text{RD}|\text{Baseline}) - cP(\text{NR}|\text{Baseline}). \quad (1)$$

Suppose the driver is given an RTI with LoA 5. Let  $P(\text{RD}|\text{LoA } 5)$  denote the probability that she resumes driving based on the RTI, and  $P(\text{NR}|\text{LoA } 5)$  be the probability that she gives no response and thus the automation tries to control the vehicle within its power, where  $P(\text{RD}|\text{LoA } 5) + P(\text{NR}|\text{LoA } 5) = 1$ , since the driver's response to the RTI is either RD or NR. By letting  $b$  denote the benefit of the fallback by the automation, we have:

$$U(\text{LoA } 5) = aP(\text{RD}|\text{LoA } 5) + bP(\text{NR}|\text{LoA } 5). \quad (2)$$

If an RTI corresponds to LoA 6, the driver may veto when she is not willing to resume driving because of her lack of confidence in taking care of the situation. Let  $P(\text{VT}|\text{LoA } 6)$  denote the probability that she vetoes (VT) the RTI. By distinguishing RD, VT and NR for her response to the RTI with LOA 6, we have:

$$U(\text{LoA } 6) = aP(\text{RD}|\text{LoA } 6) + bP(\text{VT}|\text{LoA } 6) - cP(\text{NR}|\text{LoA } 6), \quad (3)$$

where  $P(\text{RD}|\text{LoA } 6) + P(\text{VT}|\text{LoA } 6) + P(\text{NR}|\text{LoA } 6) = 1$ . Note that the values of  $a$  and  $b$  are the same in both (2) and (3).

For an RTI with LOA 6.5, the driver's response to the RTI is either RD or NR. Then we have:

$$U(\text{LoA } 6.5) = \underline{a}P(\text{RD}|\text{LoA } 6.5) - cP(\text{NR}|\text{LoA } 6.5), \quad (4)$$

where  $P(\text{RD}|\text{LoA } 6.5) + P(\text{NR}|\text{LoA } 6.5) = 1$ , and  $\underline{a}$  denotes the benefit of the driver's fallback, the quality of which might be poor (i.e.,  $\underline{a} < a$ ) because the takeover time is extremely short. Note that the value of  $c$  is the same in both (3) and (4).

### 3.2 Order relations among design alternatives for RTI

There are no differences among {Baseline, LoA 5, LoA 6} as design alternatives for an RTI message from the viewpoint of information and time allowance given to the driver, which means that  $P(\text{RD}|\text{Baseline}) = P(\text{RD}|\text{LoA } 5) = P(\text{RD}|\text{LoA } 6)$ . In the case of an RTI with LoA 6.5, the takeover time given to the driver is far shorter than any one of the above three cases, e.g.,  $P(\text{RD}|\text{LoA } 6.5) < P(\text{RD}|\text{Baseline})$ .

Thus, we have:

$$\begin{aligned} P(\text{NR}|\text{LoA } 5) &= P(\text{VT}|\text{LoA } 6) + P(\text{NR}|\text{LoA } 6) \\ &= P(\text{NR}|\text{Baseline}) < P(\text{NR}|\text{LoA } 6.5). \end{aligned} \quad (5)$$

It is then straightforward to have the following order relation:

$$U(\text{LoA } 6.5) < U(\text{Baseline}) < U(\text{LoA } 6) < U(\text{LoA } 5). \quad (6)$$

### 3.3 Is SAE conditional driving automation a sensible target to be aimed at?

What the order relation (6) implies may be summarized in the following six points.

(A) What is called "Baseline" in this paper is the RTI design that tells the driver simply to "Intervene and resume driving within  $T$  seconds time." The automation disengages either when the driver resumes driving or when she gives no response for  $T$  seconds. The Baseline is the RTI message that is suggested in (SAE 2016). The order relation (6) shows that there exist design alternatives for RTI messages that are better than the Baseline. In other words, the Baseline is not a reasonable target to translate into reality.

(B) Suppose that the automation issues an RTI of Baseline type and the driver dithers over whether to intervene or not due to her lack of confidence in taking care of the



situation. The only action she can take in that case may be to give no response to the RTI for more than  $T$  seconds. The consequence would be that neither the driver nor the automation controls the vehicle. If the RTI were designed as the one with LoA 6, the driver can express explicitly that she cannot take over control by vetoing the RTI. Then the automation would lead the vehicle to a minimal risk condition. The quantity defined as  $[U(\text{LoA } 6) - U(\text{Baseline})]$  may be defined as the *value of veto power*.

(C) An optimal design among {Baseline, LoA 5, LoA 6, and LoA 6.5} is an RTI with LoA 5 that tells the driver to “Intervene and resume driving within  $T$  seconds time.” In this case the system disengages when it confirms that the driver has started vehicle operation. If the driver fails to respond to the RTI, the system continues vehicle operation within its power to act with minimal risk. Note that this type of RTI with LoA 5 is outside the reach of LoDA 3 in (SAE 2016) which assumes that the DDT fallback should be performed by the driver, not by the automation.

(D) Is there any possibility that the RTI with LoA 5 may be regarded as identical to LoDA 4 in Table 1? The answer is negative. It is defined in (SAE 2016) that LoDA 4 must be able to perform DDT fallback without any expectation that the driver will respond to an RTI, and that the driver does not need to be receptive and respond to an RTI. Such a situation does not match what we discussed in Sect. 3.

(E) The facts pointed out so far indicate that Table 1 gives an incomplete list of LoDAs because an important level is missing between LoDA 3 and LoDA 4. Note that a candidate for filling in the missing level might be *High Automation* defined in (SAE 2014), the previous version of SAE J 3016 mentioned in (SAE 2016). It is defined that High Automation “will alert a human driver several seconds in advance of the need to resume the dynamic driving task (i.e., by issuing a request to intervene); however, the automated driving system is capable of restoring the vehicle to a minimal risk condition automatically if a human driver fails to resume the dynamic driving task when prompted” (SAE 2014). The High Automation in (SAE 2014) is different from LoDA 4 in (SAE 2016).

(F) There are two ways to solve the missing level problem. One is to place the SAE High Automation in (SAE 2014) between LoDA 3 and LoDA 4 in (SAE 2016). The other is to revise the definition of LoDA 3 so that the automation may perform DDT fallback when the driver fails to respond to the RTI. Otherwise, LoDA 3 in (SAE 2016) cannot be a technically sound target for development.

## 4 Request to intervene accompanied by automatic safety control

Suppose that LoDA 3 in SAE (2016) determines that the operational design domain limits are about to be exceeded or that there is a DDT performance-relevant system failure. The automation then issues an RTI and waits for the driver’s response. During that time, the vehicle approaches continually toward a critical point. However, if the automation applies some automatic safety control action (such as partial braking) when an RTI is issued, the following benefits might be expected: (i) the automatic safety control action can be a trigger for the driver to sense situational changes and may improve her receptivity and response to the RTI, and (ii) the automatic safety control action can make the time to the critical point longer and may improve the driver’s performance by reducing her sense of panic.

This section investigates whether such automatic safety control is effective or not in a machine-initiated transfer of authority from the automation to the driver.

### 4.1 Expected utility for an RTI accompanied by an automatic safety control action

Let  $P(\text{RD}|\text{Baseline}, \text{SC})$  denote the probability that the driver resumes driving based on an RTI of Baseline type which is accompanied by automatic safety control (SC), and  $P(\text{NR}|\text{Baseline}, \text{SC})$  be the probability that no response is given to the RTI, where  $P(\text{RD}|\text{Baseline}, \text{SC}) + P(\text{NR}|\text{Baseline}, \text{SC}) = 1$ , since the driver’s response to the RTI is either RD or NR. Then the expected utility  $U(\text{Baseline}, \text{SC})$  for an RTI of Baseline type is given by:

$$U(\text{Baseline}, \text{SC}) = aP(\text{RD}|\text{Baseline}, \text{SC}) - cP(\text{NR}|\text{Baseline}, \text{SC}). \quad (7)$$

For an RTI with LoA 5 in which the driver’s response is either RD or NR, we have:

$$U(\text{LoA } 5, \text{SC}) = aP(\text{RD}|\text{LoA } 5, \text{SC}) + bP(\text{NR}|\text{LoA } 5, \text{SC}), \quad (8)$$

where  $P(\text{RD}|\text{LoA } 5, \text{SC}) + P(\text{NR}|\text{LoA } 5, \text{SC}) = 1$ .

In case of an RTI with LoA 6, we have:

$$U(\text{LoA } 6, \text{SC}) = aP(\text{RD}|\text{LoA } 6, \text{SC}) + bP(\text{VT}|\text{LoA } 6, \text{SC}) - cP(\text{NR}|\text{LoA } 6, \text{SC}), \quad (9)$$

where  $P(\text{RD}|\text{LoA } 6, \text{SC}) + P(\text{VT}|\text{LoA } 6, \text{SC}) + P(\text{NR}|\text{LoA } 6, \text{SC}) = 1$ .

For an RTI with LoA 6.5, we have:

$$U(\text{LoA } 6.5, \text{SC}) = aP(\text{RD}|\text{LoA } 6.5, \text{SC}) - cP(\text{NR}|\text{LoA } 6.5, \text{SC}), \quad (10)$$

where  $P(\text{RD}|\text{LoA } 6.5, \text{SC}) + P(\text{NR}|\text{LoA } 6.5, \text{SC}) = 1$ .

## 4.2 Effects of automatic safety control on design of RTI

It is natural to expect that automatic safety control would be effective in reducing the driver's sense of panic, thus increasing chances for her to reach her own decision within  $T$  seconds time. As a matter of fact, if the vehicle slows down automatically when the automation issues an RTI, the driver might be a bit more relaxed to take over control, compared to the case in which the vehicle continues to move fast. To express such a situation, the following inequalities are introduced:  $P(RD|Baseline, SC) > P(RD|Baseline)$  for an RTI of the Baseline type,  $P(RD|LoA 5, SC) > P(RD|LoA 5)$  for RTI with LoA 5, and  $P(RD|LoA 6, SC) > P(RD|LoA 6)$ , as well as  $P(VT|LoA 6, SC) > P(VT|LoA 6)$  for RTI with LoA 6. However, in case of RTI with LoA 6.5, the automatic safety control might not be effective in reducing the driver's sense of panic because the automation disengages immediately, namely  $P(RD|LoA 6.5, SC) = P(RD|LoA 6.5)$ .

It is easy to see that an automatic safety control is effective when coupled with an RTI of Baseline type. Since  $P(RD|Baseline) + P(NR|Baseline) = 1$  and  $P(RD|Baseline, SC) + P(NR|Baseline, SC) = 1$ , substitutions will yield:

$$U(Baseline, SC) - U(Baseline) = (a + c) \{P(RD|Baseline, SC) - P(RD|Baseline)\} > 0. \quad (11)$$

In a similar manner, we have the following result in case of an RTI with LoA 6:

$$U(LoA 6, SC) - U(LoA 6) = (a + c) \{P(RD|LoA 6, SC) - P(RD|LoA 6)\} + (b + c) \{P(VT|LoA 6, SC) - P(VT|LoA 6)\} > 0. \quad (12)$$

However, an RTI with LoA 6.5 can gain no benefit from automatic safety control. There is none because of the immediate disengagement of the automation after issuing an RTI. Therefore,

$$U(LoA 6.5, SC) - U(LoA 6.5) = (a + c) \{P(RD|LoA 6.5, SC) - P(RD|LoA 6.5)\} = 0. \quad (13)$$

In the case of an RTI with LoA 5, whether automatic safety control is effective varies depending on the sign of  $(a - b)$ , because:

$$U(LoA 5, SC) - U(LoA 5) = (a - b) \{P(RD|LoA 5, SC) - P(RD|LoA 5)\}. \quad (14)$$

This may be rephrased as follows: (a) If the driver is more dependable than the automation in emergency situations, the coupling of an RTI with LoA 5 and automatic safety control is beneficial. (b) However, if the driver is less dependable

than the automation in emergency situations, the automatic safety control can make a situation worse because it may cause the driver to perform poor fallback in emergencies.

This section has restricted the discussions only to the case in which the automatic safety control may be effective in reducing the driver's sense of panic. However, the automatic safety control may be also effective to make the driver's fallback maneuvers smoother. In such a case, the coefficient  $a$  may be replaced by  $a^*$  where  $a^* > a$ . On the other hand, the coefficient  $c$  may be replaced by  $c^*$  in which  $c^* < c$ , because the cost of a slower vehicle's accident would be milder than that of a faster vehicle. Under this kind of assumption, similar results as (11)–(14) can be obtained but in a more favorable way to the automatic safety control. In order to avoid giving almost similar equations, this paper suppressed the discussion of the case where the automatic safety control works to make the driver's fallback maneuvers smoother.

## 4.3 Automatic safety control maintains the order relations among RTI design alternatives

Although automatic safety control can have effects on each of RTI design differently, it maintains the same order relation as (6):

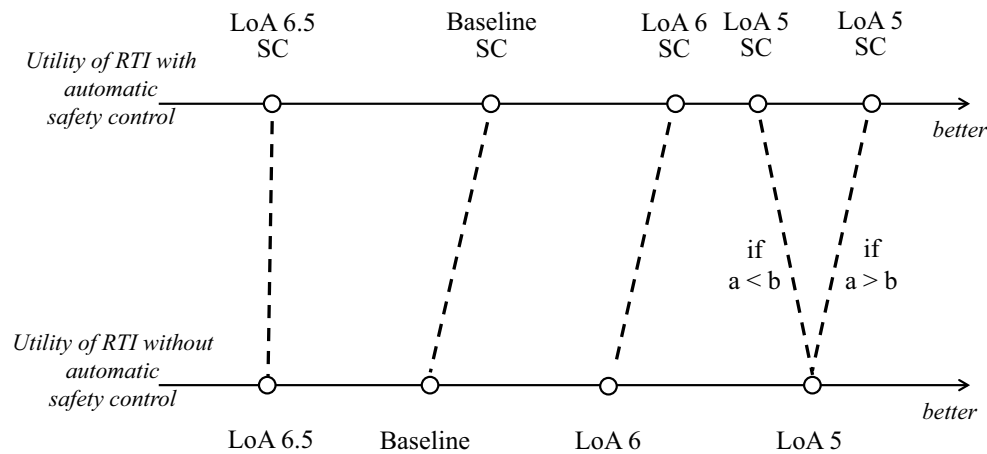
$$U(LoA 6.5, SC) < U(Baseline, SC) < U(LoA 6, SC) < U(LoA 5, SC). \quad (15)$$

The order relation (15) confirms that the points made in Sect. 3.3 are still valid. Figure 1 gives a summary of order relations among design alternatives for RTI messages, as well as effects of automatic safety control on the utility of each RTI design.

## 5 Concluding remarks

There are various ways to allocate DDT between the driver and the automation. Table 1 shows SAE's five types of function allocation schema, ranging from LoDA 1 to LoDA 5. Among them, LoDA 3 is a dynamic function allocation in which some DDT tasks are transferred from the automation to the driver based on the decision of the automation. When necessary, the automation issues an RTI to the driver with an expectation that she notices it and accepts the trading of authority for vehicle control from the automation to her. Machine-initiated trading of authority from the automation to the human is not easy though, even for cases of professional human operators, as the following example from aviation suggests.

**Example 5** Air France Flight 447 crashed into the Atlantic Ocean in June 2009. While flying at an altitude of 35,000 feet in an area of turbulence, the measured airspeed dropped



**Fig. 1** Order relation among design alternatives for an RTI message

suddenly due to the obstruction of the Pitot probe by ice crystals. That caused disconnection of autopilot and auto-thrust. The human pilot tried to control the aircraft in this unexpected event. However, he sometimes gave over-correcting input to the aircraft. The aircraft finally stalled and did not recover (BEA 2011).

Sections 3 and 4 in this paper suggest that similar problems in trading of authority for highway vehicle control can happen in LoDA 3. The difficulty could be more severe than in the case of aircraft, because most automobile drivers are non-professionals and their knowledge/skills for using automation may not be adequate.

This paper has proven mathematically that LoDA 3 may not be a meaningful target to be translated into reality, because it is risky to hand over control authority to a driver who has not been actively involved in DDT while the automation is engaged. It is claimed in this paper that a reasonable way to put LoDA 3 into reality is to revise its definition so that automation can continue vehicle control when no RTI response is given by the driver in a timely manner. The resulting definition of LoDA 3 would be almost equivalent to that of the High Automation given in SAE (2014).

Another possible and promising alternative to make LoDA 3 sufficiently attractive may be to use the *haptic shared control* in the phase of trading of authority for vehicle control from the automation to the driver in the situation where the automation issued an RTI. The concept of haptic shared control itself is broad, and there are many application domains (see, e.g., Abbink et al. 2012 and Flemisch et al. 2012 for overviews). Haptic shared control allows the human operator to communicate with the automation through forces on the control interface. An important concept in haptic

shared control is the level of haptic authority (LoHA), which describes “how forceful the human-automation interface connects the human input to automation inputs” (Abbink et al. 2012). By selecting an appropriate LoHA, a new level of automation can be obtained in the scale ranging from full automation to full manual control. Moreover, LoHA may be changeable dynamically either in the machine-initiated manner or in the human-initiated manner (Abbink et al. 2012; Flemisch et al. 2012).

By noting that the driver may not be good at taking on authority from the automation after enjoying automated driving over a period of time, we have to ask whether it is possible to let the automation collaborate with the driver during the process of trading of authority for vehicle control. A natural answer to the question may be to apply the LoHA concept to the authority trading process and change it flexibly and dynamically depending on the situation when trading of authority for vehicle control is in progress from the automation to the driver. Wada and his colleagues have proposed a “shared authority mode that connects the automated and human driving modes” for LoDA 3 (Wada et al. 2016; Wada and Kondo 2017). There are similar research topics, such as, “Can a machine-initiated method for variable authority cause conflict of intentions, automation surprises, or distrust of automation?” or “Can a human-initiated method increase the driver’s burden?” or “Are there any tradeoffs when we use just a portion of possible ranges for LoHA?” These questions are closely related to issues of machine-initiated trading of authority in *adaptive automation* (Inagaki 2003; Inagaki and Sheridan 2012; Inagaki et al. 2007) and human-initiated trading of authority in *adaptable automation* (Opperman 1994; Scerbo 2001). Extensive studies are necessary to help the driver in the phase of trading of authority upon RTI.



**Acknowledgement** This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (S) 15H05716.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Abbink DA, Mulder M, Boer ER (2012) Haptic shared control: smoothly shifting control authority? *Cogn Technol Work* 14(1):19–28
- BASt (2013) Legal consequences of an increase in vehicle automation: final report of the project group. BASt-Report F83
- BEA (2011) On the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro—Paris, Interim Report no. 3
- Billings CE (1997) Aviation automation—the search for a human-centered approach. LEA, Mahwah
- Boeing (2005) 777 flight crew operation manual. The Boeing Company, Chicago
- Botargues P (2009) Airbus new auto pilot/flight director TCAS mode—enhancing flight safety during TCAS manoeuvres. *Flight Airworth Support Technol (FAST)* Airbus Tech Mag 45:10–15
- Endsley MR, Kaber DB (1999) Levels of automation and effects on performance: situation awareness and workload in a dynamic control task. *Ergonomics* 42(3):462–492
- Flemisch F, Hessen M, Hesse T, Kelsch J, Schieben A, Beller J (2012) Towards a dynamic balance between humans and automation: authority, ability, responsibility and control in shared and cooperative control situations. *Cogn Technol Work* 14(1):3–18
- Gasser TM, Westhoff D (2012) BASt-study: definitions of automation and legal issues in Germany. 2012 Road Vehicle Automation Workshop, Transportation Research Board, July 25, 2012
- Gold C, Damböck D, Lorenz L, Bengler K (2013) “Take over!” how long does it take to get the driver back into the loop?. In: *Proceedings of the human factors and ergonomics society 57th annual meeting*, pp 1938–1942
- Happee R, Gold C, Radlmayr J, Hergeth S, Bengler K (2017) Take-over performance in evasive manoeuvres. *Accid Anal Prev* 106:212–222
- Inagaki T (2003) Adaptive automation: sharing and trading of control. In: Hollnagel E (ed) *Handbook of cognitive task design*. Erlbaum, Mahwah, pp 147–169
- Inagaki T (2005) Situation-adaptive responsibility allocation for human-centered automation. *Trans Soc Instrum Control Eng Jpn* 31(3):292–298
- Inagaki T, Inoue K (1993) Adaptive choice of a safety management scheme upon an alarm under supervisory control of a large-complex system. *Reliab Eng Syst Saf* 31(1):81–87
- Inagaki T, Sheridan TB (2012) Authority and responsibility in human-machine systems: probability theoretic validation of machine-initiated trading of authority. *Cogn Technol Work* 14(1):29–37
- Inagaki T, Moray N, Itoh M (1998) Trust, self-confidence, and authority in human-machine systems. In: *Proceedings IFAC symposium on man-machine systems*, pp 431–436
- Inagaki T, Itoh M, Nagai Y (2007) Support by warning or by action: which is appropriate under mismatches between driver intent and traffic conditions? *IEICE Trans Fundam* 90-A(11):2540–2545
- Jou Y-T, Yenn T-C, Lin CJ, Yang C-W, Chiang C-C (2009) Evaluation of operators’ mental workload of human-system interface automation in the advanced nuclear power plants. *Nucl Eng Des* 239:2537–2542
- Kaber DB, Endsley MR (2007) The effects of levels of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theor Issues Ergon Sci* 5:113–153
- Kaminski-Morrow D (2009) XWB could auto descend in cabin pressure emergency. *Flight Int* 18–24 August 2009
- Lindstrom V, Winroth M (2010) Aligning manufacturing strategy and levels of automation: a case study. *J Eng Technol Manag* 27:148–159
- Lu Z, Happee R, Cabral CDD, Kyriakidis M, de Winder JCF (2016) Human factors of transitions in automated driving: a general framework and literature survey. *Transp Res Part F* 43:183–198
- Merat N, Hamish Jamson A, Lai FCH, Daly M, Carsten OMJ (2014) Transition to manual: driver behavior when resuming control from a highly automated vehicle. *Transp Res Part F* 27:274–282
- Moray N, Inagaki T, Itoh M (2000) Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *J Exp Psychol Appl* 6(1):44–58
- NHTSA (2013) Preliminary statement of policy concerning automated vehicles. US National Highway Traffic Safety Administration, May 30, 2013
- NHTSA (2016) Federal automated vehicles policy—accelerating the next revolution in roadway safety. US National Highway Traffic Safety Administration, September 30, 2016
- Opperman R (1994) Adaptive user support. Erlbaum, Hillsdale
- Pacaux M-P, Debernard S, Godin A, Rajaonah B, Anceaux F, Vanderhaegen F (2011) Levels of automation and human-machine cooperation: application to human–robot interaction. In: *Proceedings 18th IFAC world congress*, pp 6484–6492
- Parasuraman R, Sheridan TB, Wickens CD (2000) A model for types and levels of human interaction with automation. *IEEE Trans Syst Man Cybern A Syst Humans* 30(3):286–297
- Petermeijer SM, Cieler S, de Winter JCF (2017) Comparison spatially static and dynamic vibrotactile take-over requests in the driver seat. *Accid Anal Prev* 99:218–227
- SAE (2014). Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems (Standard No. J3016), SAE International, January 2014
- SAE (2016) Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles (Surface Vehicle Recommended Practice: Superseding J3016 Jan 2014), SAE International, September 2016
- Scerbo M (2001) Adaptive automation. In: Karwowski W (ed) *International encyclopedia of ergonomics and human factors*. Taylor & Francis, London, pp 1077–1079
- Sheridan TB (1992) *Telerobotics, automation, and human supervisory control*. MIT Press, Cambridge
- Sheridan TB (1999) Human supervisory control. In: Sage AP, Rouse WB (eds) *Handbook of systems engineering and management*. Wiley, New York, pp 592–628
- Sheridan TB (2002) Humans and automation: system design and research issues. *Human Factors & Ergonomics Society & Wiley*, New York
- Sheridan TB, Verplank WL (1978) Human and computer control of undersea teleoperators (Tech. Rep.). Man-Machine Systems Laboratory, Department of Mechanical Engineering, MIT, Cambridge, MA
- Wada T, Kondo R (2017) Shared authority mode: Connecting automated and manual driving for smooth authority transfer. In: *Proceedings of FAST-zero 2017*
- Wada T, Sonoda K, Okasaka T, Saito T (2016) Authority transfer method from automated to manual driving via haptic shared

- control. In: Proceedings of IEEE international conference on systems, man, and cybernetics 2016
- Woods DD (1989) The effects of automation on human's role: experience from non-aviation industries. In: Norman DA, Orlady H (eds) Flight deck automation: promises and realities, NASA CP-10036, Moffet Field, CA, pp 61–85
- Zeeb K, Buchuner A, Schrauf M (2015) What determines the take-over time? An integrated model approach of driver take-over after automated driving. *Accid Anal Prev* 78:212–221
- Zeeb K, Buchuner A, Schrauf M (2016) Is take-over time all matters? The impact of visual-cognitive load on driver take-over quality after conditionally automated driving. *Accid Anal Prev* 92:230–239