Necessity of supporting situation understanding to prevent over-trust in automation

ITOH Makoto

Department of Risk Engineering, University of Tsukuba, Tsukuba 305-8573, Japan(itoh@risk.tsukuba.ac.jp)

Abstract: It is necessary to clarify how a human operator comes to expect that an automation will perform a task successfully even beyond the limit of the automation. This paper proposes a way of modeling trust in automation by which it is possible to discuss how an operator's trust in automation becomes over-trust. On the basis of the model, an experiment using a micro world was conducted to examine whether the range of user's expectation *per se* surpasses the limit of the capability of the automation. The results revealed that informing the human operator of the functional limit of capability of automation without giving an appropriate reason was effective but not perfect for avoiding human operator's over trust. It was also shown that the understanding of the automation's limitation can be changed through experiences due to confusion about the situation, therefore, it is necessary to support adequate situation understanding of the human operator in order to prevent over-trust in automation.

Keyword: automation, trust, situation awareness, human-machine interaction

1 Introduction

Computerized machines are becoming intelligent and powerful. Human operators may become too reliant on such intelligent machines, because they appear to perform tasks in a fully autonomous manner. However, the machines' capabilities and/or functions are still limited. Preventing over-trust in automation is thus one of the crucially important issues in human-machine systems and the research on these systems has been given a considerable impetus.

The notion of over-trust in automation is very complicated [1]. Therefore, it is important to distinguish over-trust from 'complacency'. Even though discussions on defining the term ‘complacency’ have not come to a mutual consensus (*e.g.* see Parasuraman, Sheridan, and Dickens [2]), the term, in layman's term refers to ‘lack of vigilance’ [3]. According to Singh, Molloy and Parasuraman [4], the tendency to be complacent is dependent on the human operator's attitude. Conversely, even a vigilant human operator may still trust automation excessively if he or she misunderstands what the automation can do. If the misunderstanding comes not from the human operator's attitude but from some environmental factors, such over-trust may occur even when an operator is highly motivated and trained. Ockerman and Pritchett [5] discussed professional workers' over-reliance on task guidance systems.

In general, over-trust in automation is a special case of the trust in automation. In order to clarify how to prevent over-trust, it is necessary for us to understand how appropriate trust can become inappropriately high. As Parasuraman and Riley [6] pointed out, there are myriad factors that affect human trust in automation.

There are various definitions of trust in automation. The detailed discussion can be found in Lee & See [7]. Among them, Muir's [8] and Lee & Moray's definitions [9] are essential. According to Muir[8], trust has three dimensions: *predictability*, *dependability*, and *faith*. Lee & Moray[9] expanded Muir's model and proposed the following four dimensions: *Purpose*, *Process*, *Performance*, *and Foundation*. However, there are hardly any explanations given about over-trust in the aforementioned models.

This paper proposes a model of trust in automation within which it is possible to discuss how operator's trust in automation becomes over-trust. On the basis of the model of trust, a cognitive experiment using a micro world was conducted in order to investigate how a human operator comes to expect that an automated system can perform a task successfully even

Received date: December 8, 2009
(Revised: April 20,2010)

beyond its limit of automation. The results show that the possibility of operators' over-trust was reduced if the operators were informed of the reason for the limit of the capability of automation. However, providing such information was not sufficient to prevent over-trust. A comprehensive analysis of the experimental data revealed that a misunderstanding of the task situation had caused some of the over-trust. The results suggest that the support for human operators towards understanding the situation correctly is indeed pivotal to preventing over-trust.

2 Over-trust

2.1 Trust and Over-Trust

According to Lee & Moray's definition [9], three types of over-trust in automation are delineated: over-trust in purpose, over-trust in process, and over-trust in performance.

The differences existing among the three can be understood by looking at the following examples.

Example 1: SRS airbags

SRS airbags became prevalent in Japan during the 1990s and today, most cars in Japan are equipped with airbags. However, there were many problems associated with the use of SRS airbags. For instance, drivers tended to rely on SRS airbags instead of seat belts [10]. There were several cases in which a driver was killed or seriously injured by an SRS airbag when the vehicle crashed into something. In such cases, the drivers had not fastened their seat belt at the time of the accident. One possible reason for the seat belt non-use was that many drivers regarded the SRS airbag as an alternative to a seat belt. However, the drivers' understanding was inappropriate –an SRS airbag is a restraint system that is supplemental to a seat belt. In view of this, such driver's inappropriate reliance on an SRS airbag can be deemed as "over-trust in purpose" in the systems. Conversely, there were complaints from some drivers about the non-activation of the SRS airbags during traffic collision. The reason for the non-activation of the SRS airbags, in most cases was not due to their malfunction but rather because the situation was beyond the system's operative condition. Thus, it can be said that the complaint is due to the drivers' "over-trust in process" in the systems.

Example 2: Adaptive cruise control systems

An Adaptive Cruise Control (ACC) system is a partial automation for longitudinal control, designed to reduce the driver's workload by freeing him or her from frequent pedal manipulation. An ACC system maintains vehicle speed at a driver-specified level when there is no slower vehicle ahead (*i.e.* the preceding vehicle) and maintains headway distance at a driver-specified distance if necessary (Fig. 1).

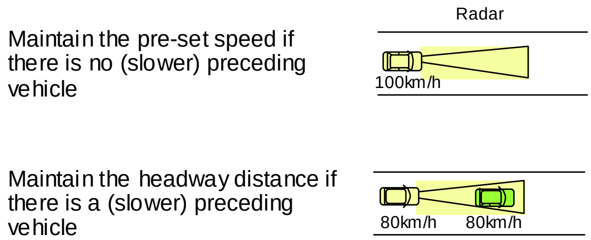

Fig. 1 Adaptive Cruise Control System.

There are multiple types of over-trust in ACC systems. Two typical examples are shown below (Fig. 2).

2.2.1 Over-trust in purpose

What happens if there is a fixed obstacle (including a stopped car) ahead when an ACC system is maintaining the vehicle speed (*i.e.* there is no vehicle ahead)? Since ACC systems are designed to follow a moving object but not to avoid a crash into a fixed object, a fixed obstacle is ignored by the ACC systems. If the driver does nothing or delays in hitting the brake or rotating the steering wheel expecting that the ACC system will stop the vehicle (or at least reduce the vehicle speed) before touching the obstacle, a collision will happen. The driver's inappropriate expectation is an example of "over-trust in purpose" in an ACC system [11].

2.2.2 Over-trust in performance

This is another type of over-trust. When an ACC system detects the deceleration of the target vehicle, it slows down the host vehicle at some deceleration rate. Note here that the maximum deceleration rate which can be affected by the ACC system is limited to a lower level, say, 0.25G. When the lead vehicle makes a rapid deceleration, say at 0.4G, the ACC's application of the brakes may not be powerful enough to avoid a collision. If the driver relies on the ACC system by expecting that the ACC system can avoid a collision even when the deceleration of the preceding vehicle is high, this can be regarded as "over-trust in performance" in the ACC system [12].

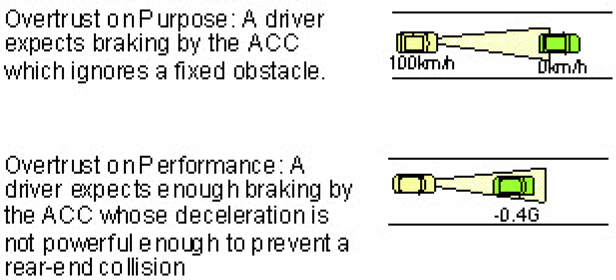

Fig. 2 Two types of over-trust.

2.2 Structure of trust in performance

This paper focuses on trust in performance and proposes a way of modeling trust in automation through which it is possible to discuss how an operator's trust in automation becomes over-trust. The horizontal axis in Fig.3 represents the level of difficulty for an automated system (LDA) to perform a given task. The vertical axis represents the reliability of the automation at the associated task condition. It is assumed that there exists a functional limit within which the automation works successfully —(actual automation range or Actual-AR). However, operations are often restricted to an easier range of situations than the Actual-AR. Thus, it is assumed that the second limit (assured automation range or Assured-AR) is set to guarantee the automation works correctly.

This paper proposes a structure of trust in performance as shown in Fig. 4. This model is based on Muir's model [8]. The key components are: Faith, Dependability, and Predictability. Faith (F) is regarded as the range within which a human operator wants the automation to work as the operator wishes.

The vertical axis in Fig. 4 represents the operator's willingness level to rely on the automation. The willingness level is assumed to range from 0 (complete distrust) to 1 (complete trust). The dotted line in Fig. 4 is a trust-curve which represents the human operator's willingness to rely on the automation as a function of LDA.

As shown in Fig. 4 below, the area of F can be divided into three sub-areas: D (dependability), UD (undependability), and UP (unpredictability). A human operator feels that the automation is reliable and dependable when the situation is in D on the basis of his or her past experiences. On the other hand, the operator feels the automation to be untrustworthy when the situation is in UD, based on his or her experiences. The behavior of the automated system in both D and UD are "predictable" for a human operator. There exist some unpredictable conditions (UP) in which a human operator is still not sure whether the automation is dependable or not.

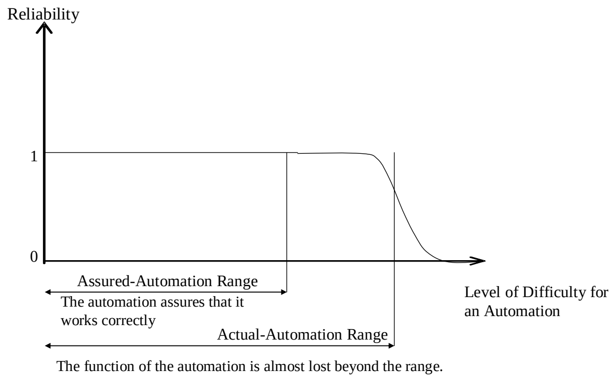

Fig. 3 Automation range and reliability.

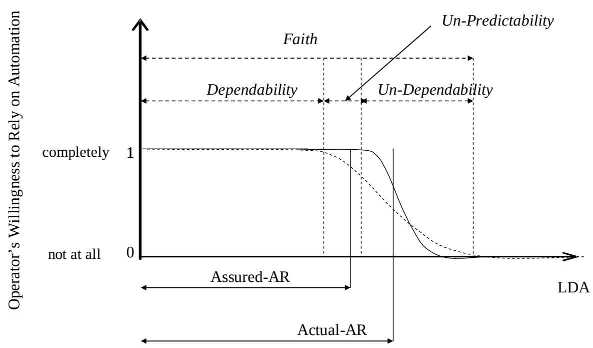

Fig. 4 Structure of trust in performance.

2.3 Over-trust in performance

Regarding trust in performance, over-trust is a kind of an inappropriate calibration in which trust exceeds the automation's capabilities. In this paper, such inappropriate calibrations are described as follows. If Actual-AR is a subset of D, it can be said that the trust is excessive (Fig. 5).

Interestingly, this model can discuss the potential of over-trust (Fig. 6). In this case, the upper bound of Faith (denoted as F) is greater than the Actual-AR.

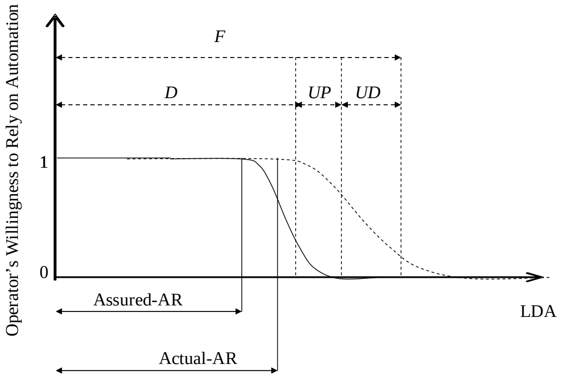

Fig. 5 Example of over-trust.

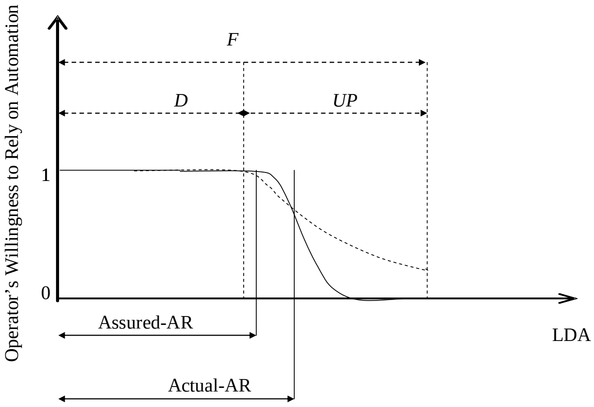

Fig. 6 Potential over-trust.

2.3 Causes of over-trust

In many accident cases, an automated system was used even though the automation was not suitable for the situation. In other words, the situation was beyond the Actual-AR.

It is assumed that human operators receive training in the use of automated systems and that the operators understand the Assured-AR. On the other hand, understanding of the Actual-AR is not necessarily adequate. There are two types of failure to understand the Actual-AR:

(1) The human operators are not explicitly informed of the functional limits of the automation.
(2) The functional limit of the automation is given to the human operator. However, the reason for the functional limit is not given. The operator may think that the 'true' limit is greater than the given functional limit.
(3) Both the functional limit of automation and the reason for it, are given to the operators. However, their understanding of the Actual-AR changes, on the basis of their experiences of using the automation is not a given.

Based on these assumptions, we then investigate how the above factors affect human operators' tendency for excessive trust by conducting an experiment.

3 Experiment

3.1 Mixed juice processing plant

In this experiment, a computer-controlled simulation of a mixed juice pasteurizing plant was used as shown in Fig.7 below [13, 14].

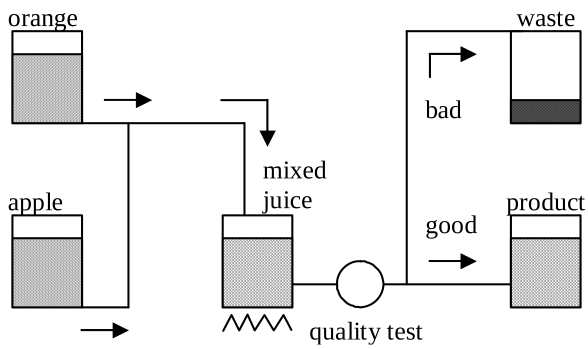

Fig. 7 Mixed juice processing plant. [13, 14]

The production process of the mixed juice is automated. However, this automated process is not always successful because the quantity of raw juice that flows into the mixture vat does not always exactly equal that specified in the order sheet. In the present paper, *supply error* (E) refers to the difference between the desired mass and the actual mass in the mixture vat. The automatic pasteurization is assumed to be successful in most cases if E is within 5% of the desired mass. However, if E > 5%, the pasteurization time should be manually recalculated according to the actual mass, otherwise the automatic pasteurization fails due to residual germs in most cases (Fig. 8). If E < 3%, it is guaranteed that the automation pasteurizes the juice successfully.

The task imposed on the operator is the supervision of the automated system. Operators are encouraged to rely on the automatic pasteurization system as much as possible, because orders to produce mixed juice must be filled as fast as possible and automatic pasteurization is faster than manual pasteurization. Only when the operator believes that the automation has not set the pasteurization time properly should the operator intervene and set an appropriate pasteurization time.

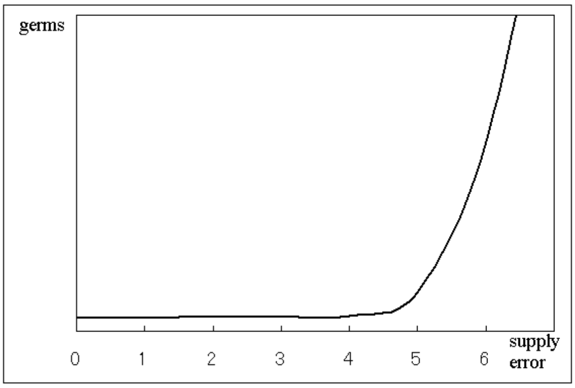

Fig. 8 Supply error and residual germs

3.2 Participants

Thirty-three undergraduate and graduate university students volunteered to participate. Volunteers received honoraria for their participation.

3.3 Design and procedure

Three types of information on the limit of automation capability are compared. Participants were randomly assigned to one of the following groups:

Group 1 (G1, participants 1a through 1k): Operators are informed that the automation will succeed in pasteurizing the juice when the supply error is less than 3%, *i.e.* only the Assured-AR is given.

Group 2 (G2, participants 2a through 2k): In addition to information given to G1, operators are informed that the automation may succeed in pasteurizing the juice when the error is less than 5%, *i.e.* both the Assured-AR and the Actual-AR are given.

Group 3 (G3, participants 3a through 3k): In addition to information given to G2, operators are informed that automation will fail to pasteurize the juice when the supply error is greater than 5% because the germs are not eliminated from the juice as shown in Fig. 8. Operators are also shown this figure, *i.e.* in addition to the Assured-AR and Actual-AR, the reason for the limit of the automation's capability is given.

The experiment which took about an hour each day, lasted three days. Participants were requested to perform 100 trials each day. On the first day, each participant was notified of the purpose and procedures of the experiment. Each participant received some training trials in order to understand when and how to intervene in the control of the process.

3.4 Measure

In each trial, the operator had to decide whether to use the automation for the pasteurization or not. Each decision on the use of automation was recorded.

4 Results and discussions

The participants were categorized into the four types shown in Table 1.

Type A: Operators who used the automation when the supply error was less than 5%.

Type B: Operators who relied completely on the automation and used it for all the 300 trials.

Type C: Operators who used the automation only when the supply error was less than 3%.

Type D: Operators who became completely reliant on the automation on the second or third day based on their experience.

Table 1 Number of participants for each type of reliance

Group	Type			
	A	B	C	D
G1	7	1	1	2
G2	7	1	1	2
G3	7	0	2	2

This categorization was done by observing the actions of the participants followed by an interview. B and C type participants had very fixed criteria to rely on the automation. The criteria seemed to be independent of the information on the automation's limitation and the task experience. This paper thus analyses only the data on the type A and D participants, in order to clarify how reliance on automation can be affected by the information on the automation's limitation and experience.

As for type A, the following two values were calculated on each day for each subject:

(1) The maximum value of the supply errors when he or she used the automation (max-auto)

(2) The minimum value of the supply errors when he or she intervened into control (min-man)

In this paper, "mode threshold" refers to the mean value of the above two. We regard the mode threshold as the operator's subjective estimate of the limit of the automation's capability. Fig. 9 depicts the trends of the mode thresholds. A two-way ANOVA on the mode threshold was conducted. The design was a 3 × 3 factorial, mapping onto Group and Day. Group was a between-operator factor, and Day was a within-operator factor. The ANOVA showed a significant main effect of Day ($F(2,36)=10.28$, $p=0.0003$), and a significant main effect of Group ($F(2,18)=10.79$, $p=0.0008$).

The main effect of Day can be interpreted as the mode threshold is increasing. Group 1 is a typical example. The main effect of Group suggests that the mode threshold is higher in G2 than in G1 and G3. In fact, the mode threshold in G2 was significantly greater than in G1 ($p=0.0008$) and G3 ($p=0.019$). These significant differences would stem from the difference in the information on the limit of the automation's capability. The participants in G2 were informed of the Actual-AR (up to 5%), thus they relied on the automation even beyond the Assured-AR (3%). However, the participants in G1 were not informed of the Actual-AR. The initial value of the mode threshold of a participant in G1 was therefore low and the value gradually increased by the day. For participants in G3, not only were they informed of the Actual-AR, but were also informed of the reason for the automation's limitation. The participants in G3 thus tended to be reluctant to rely on the automation beyond the Assured-AR. There was no significant difference between G1 and G3 by Tukey's HSD test; nevertheless, the trend of mode thresholds in G1 is different from that in G3. According to an interview held after completion of all trials, two subjects (3i and 3j) in G3 thought that the automation could be used when the supply error was less than 4.0 and 4.5 respectively even on the first day. That is, the subjective feeling of the participants 3i and 3j on the limitation of the automation capability was almost constant throughout all the days. In sum, there is a difference between G1 and G3. More of the supply error for G1 increased as compared to that of G3. Therefore, it can be claimed that operators' over reliance on automation may be avoided by informing the human operators of the functional limits of automation and the reason for the limit.

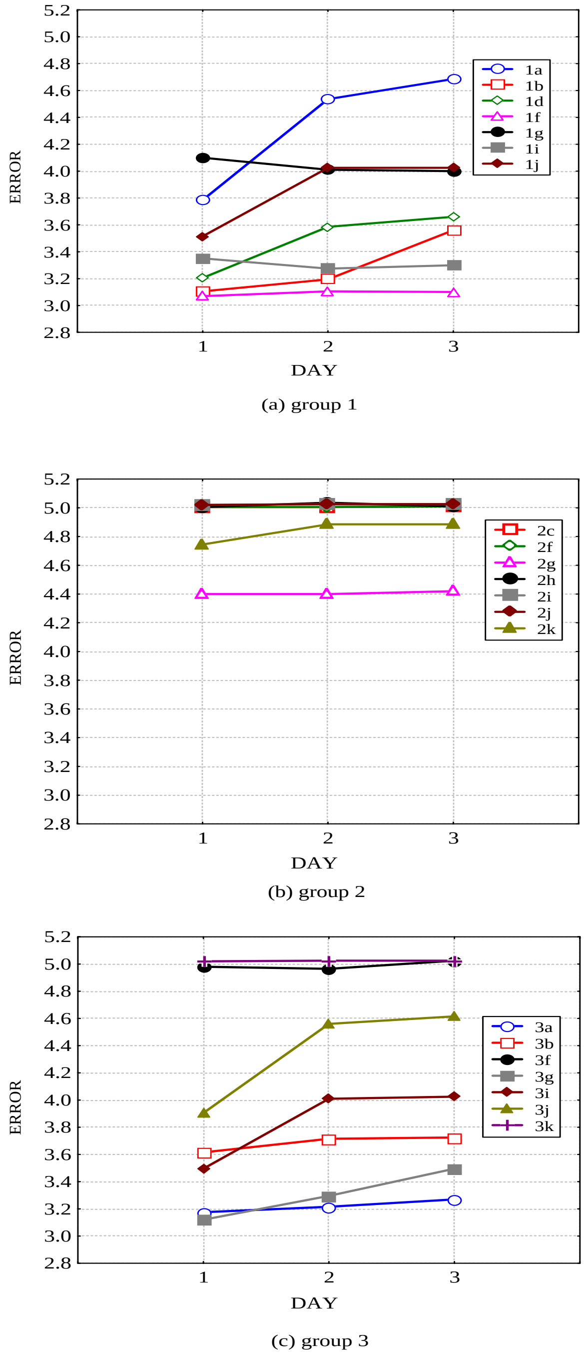

Fig. 9 Trend of mode threshold.

However, Table 1 also suggests that giving both the limits of automation and their reasons is not a perfect method of preventing over-trust in automation. Even in G3, where operators received information on the limits of automation and the reasons for them, there were two persons in type D who became completely reliant on the automation based on their experience on the second or third day.

We discuss what happened to participants in type D using participant 3h as an example. According to an interview held after completion of all the 300 trials, the participant had experience in using the automation even though the participant did not intend to use it. Because the mode thresholds were relatively high, the participant pushed the button to use the automation in most trials. Thus, the participant inadvertently pushed the button to use the automation even when the supply error was greater than 5%. The automatic pushing was successful by chance at that trial because the supply error was just slightly greater than the functional limit. This experience resulted in a change of understanding on the functional limit of the automation. A typical example of this change in mode threshold is shown in Fig.10. The horizontal axis and the vertical axis represent the trial number and the supply error at each trial, respectively. The open circles mean that the operator used the automation at the trial. On the other hand, the filled squares are trials at which the operator intervened to control the process manually. Similar phenomena were observed for participants 2a and 2d.

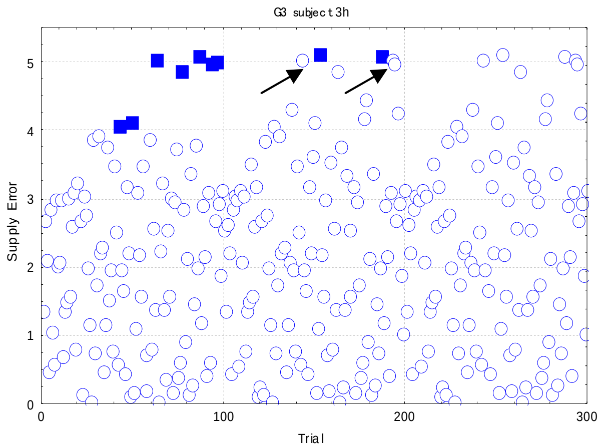

Fig. 10 Example of change of understanding of automation limit due to unintended use (Participant 3h).

5 Conclusions

The results of the experiment suggest that people may rely on automation too much if information on the functional limit of capability of automation and reasons for it are not given appropriately.

However, it is not always sufficient to inform operators of the functional limit of automation and reasons for it. Even though operators understood the limits of automation correctly, some operators changed their understanding of the automation limit based on their experiences of using automation. This phenomenon can occur in the real world. It may be difficult for operators to distinguish whether the current operating conditions are within the functional limit or not. If an operator uses the automation inadvertently when the current operating condition seems to be beyond the functional limit, the operator may change his or her understanding on the functional limit which results in over-trust. Therefore, to reduce over-trust due to unintended use of automation, it would be necessary to support adequate understanding of the current operating conditions.

The possible contribution of this work for the safety of the nuclear industry is that system designers and/or managers should be careful about the change of the operator's understanding of the automation's limitations throughout daily operations. The findings obtained from this experiment could be applied to the motivated and trained operators, because the participants who finally became too reliant on the automation did not seem overtrust-prone or complacent-prone persons on the basis of the experimenter's observations and the interview.

Acknowledgments

The author thanks Hiromasa Inahashi for his help in data collection and analyses on experiment. I thank the reviewers for their valuable and constructive comments and suggestions in revising this paper.

References

[1] INAGAKI, T., ITOH, M.: Theoretical framework for analysis and evaluation of human's over-trust in and over-reliance on advanced driver assistance systems, In: Proceedings of European Conference on Human Centered Design for Intelligent Transport Systems, Berlin, 2010 (to appear).

[2] PARASURAMAN, R., SHERIDAN, T. B., WICKENS, C. D.: Situation awareness, mental workload, and trust in automation: viable, empirically supported cognitive engineering construct, Journal Of Cognitive Engineering And Decision Making, 2008, 2(2): 140-160.

[3] MORAY, N., INAGAKI, T.: Attention and complacency, Theoretical Issues In Ergonomics Science, 2000, 1(4): 354-365.

[4] SINGH, I. L., MOLLOY, R., PARASURAMAN, P.: Automation-induced "complacency": development of the complacency-potential rating scale, The International Journal Of Aviation Psychology, 1993, 3(2): 111-122.

[5] OCKERMAN, J., PRITCHETT, A.: Reducing overreliance on task-guidance systems [A], In: Proceedings of the IEA2000/HFES2000 Congress, San Diego, 2000, 1: 1-4.

[6] PARASURAMAN, R., RILEY, V.: humans and automation: use, misuse, and abuse,, Human Factors, 1997, 39(2): 230–253.

[7] LEE, J.,SEE, K. A.: Trust in automation: designing for approproaite reliance, Human Factors, 2004, 46(1): 50-80.

[8] MUIR, B.: Trust in automation: Part I. theoretical issues in the study of trust and human internvention in automated systems, Ergonomics, 1994, 37(11): 1905-1922.

[9] LEE, J., MORAY, N.: Trust, control strategies and allocation of function in human-machine systems, Ergonomics, 1992, 3(10): 1243-1270.

[10] ITOH, M.: Human Over-Trust in Automation: Is It Myth or Truth? [A], In: IEICE Technical Report, SSS2009-11, Tokyo, 2009, 21-24 (in Japanese).

[11] ITOH, M.: contributing factors to driver's over-trust in a driving support system for workload reduction, Transactions Of The Society Of Instrument And Control Engineers, 2009, 45(11): 555-561 (in Japanese).

[12] INAGAKI, T., FURUKAWA, H., ITOH, M.: Human interaction with adaptive automation: Strategies for Trading of Control under Possibility of Over-trust and Complacency [A], In: Proceedings of 1st International Conference on Augmented Cognition 2005, Las Vegas, CD-ROM,. 10 pages.

[13] ABE, G., ITOH, M., and TANAKA, K.: Dynamics of trust in automated systems dependent on occurrence patterns of malfunctions, Transactions Of The Society Of Instrument And Control Engineers, 2000, 36(12): 1138-1144 (in Japanese).

[14] ITOH, M., INAHASHI, H., TANAKA, K.: Informing Limit of Ability and its Reasons for Reducing Overtrust in Automation [A], In: Proc. XVth Triennial Congress of the International Ergonomics Association, 2003, Seoul, 1: 141-144.