Human-Machine Collaboration for Safety and Comfort

T. Inagaki*

*Department of Risk Engineering
University of Tsukuba
Tsukuba, Japan
inagaki@risk.tsukuba.ac.jp

Abstract: In the framework of *Human-centered automation,* it is usually claimed that, "the human must have final authority over the automation." However, correctness of the statement is context-dependent, by noting that authority is usually interconnected with responsibility. This paper argues that a machine-initiated automation invocation that trades authority from human to machine may be indispensable even in the framework of human-centered automation.

Keywords: Adaptive automation, Authority and responsibility, Human-centered automation, Situation awareness

1. INTRODUCTION

Suppose we are to design a human-machine system. The design decision of assigning functions to human and machine is called *function allocation*. In spite of its importance, function allocation has not become a science yet, but still a kind of art. The traditional ways of function allocation are classified into three categories: The first category is termed *comparison allocation* or, MABA-MABA (what "men are better at" and what "machines are better at") approach. The strategies of this type compare relative capabilities of humans versus machines for each function, and they allocate the function to the most capable agent. The second type is called *leftover allocation*. The strategies of this type allocate to machines every function that can be automated, and thus human operators are assigned the leftover functions to which no automation technologies are available. The third type is *economic allocation* that tries to find an allocation ensuring economical efficiency. Even when some technology is available to automate a function, if automating the function is not cost-effective, the function is assigned to the operator. The traditional strategies described above consider "who does what." Such design decisions yield function allocations that are *static*: viz., once a function is allocated to an agent, the agent is responsible for the function at all times.

Though static function allocations are easy to implement, human operators may not be very happy with them. The leftover and the economic allocation strategies do not reflect human characteristics, and treat the operators as if they were machine elements. The resulting function allocation can be elusive for the operators, and they may have to adapt to the machines unwillingly. The comparison allocation seems to be nicer for the humans than either the economic or leftover allocations. Even when the operators are allocated only functions in which humans surpass machines, the superiority may not hold at all times and on every occasion. The above discussions imply that "who does what" design decisions are not sufficient, but "who does what and when" considerations are needed, which implies that function allocation must be dynamic.

A scheme that modifies function allocation dynamically depending on situations is called an *adaptive function allocation*, or *adaptive automation* (see, e.g., Scerbo, 1996; Inagaki, 2003). In adaptive automation, functions are reallocated to humans and machines in response to changes in situations or human performance, which means that an active agent for a function changes alternately from time to time. Who is supposed to make decisions concerning when and how function allocation must be altered? The humans, or machine intelligence? Strategies for trading of authority for functions are classified into two disjoint groups; viz., human-initiated strategies and machine-initiated strategies (Scerbo, 1996). Humans are usually assumed to be responsible for the safety of the human-machine systems, and thus are considered to be in command in those systems (see, e.g., Billings, 1997). Does this mean that the human-initiated strategies are the only solution for trading of authority?

A human can fail to give a proper directive to the machine in various ways. One of most obvious cases may be where the human's *situation awareness* (SA) is inappropriate or incomplete. When a human's SA is in appropriate or incomplete, the human's decision and action are likely to be incorrect. It is noted that correct SA by the human does not assure that an unwanted result can be avoided, especially when allowable time is limited for the human to take a necessary countermeasure or to give a directive to the machine. Today's machines can sense and analyze a situation, decide what must be done, and implement control actions. Should such an intelligent machine do nothing if it is not given a directive by a human, even when it has detected that the human is late in taking a control action that is needed in a given situation? Should such an intelligent machine sit back when it detects a human's apparently inappropriate control action, by assuming that the human must have some good reason for doing so? Allowing a machine to take a corrective control action when it believes that the human is late in taking a necessary measure or behaving inappropriately implies that the authority is traded to the machine temporarily based on the judgment of the machine (Inagaki, 2006).

This paper argues that a machine-initiated automation invocation may be indispensable for assuring safety of human-machine systems, even in the framework of human-centered automation.

2. HUMAN-COMPUTER INTERACTIONS

Humans perceive the situation, decide what must be done, and implement a control action. In the design of artifacts to assist humans, it is useful to distinguish the following four classes of functions: (1) Information acquisition, (2) Information analysis, (3) Decision and action selection, and (4) Action implementation.

Example: Traffic alert and collision avoidance system (TCAS) is a family of airborne devices designed to help pilots to avoid a mid-air collision (US Dept. of Transportation & FAA, 2000). Its functionalities are described as follows.

(1) Information acquisition: TCAS sends interrogations at 1030 MHz that transponders on nearby aircraft respond to at 1090 MHz. By decoding the replies, the position and altitude of the nearby aircraft can be known.

(2) Information analysis: Based on the range, altitude, and bearing of nearby aircraft, TCAS performs *range and altitude tests* to determine whether the aircraft is a *threat* or not.

(3) Decision and action selection: When the nearby aircraft is declared a threat, TCAS selects an avoidance maneuver (to climb or to descend) that will provide adequate vertical miss distance from the threat. If the threat aircraft is equipped with TCAS, the avoidance maneuver will be coordinated with the threat aircraft.

(4) Action implementation: TCAS issues the *resolution advisory* (RA) to let the pilot know the appropriate avoidance maneuver. However, TCAS does not perform any avoidance maneuver itself. It is the human pilot who implements the avoidance maneuver.

In the above example, information acquisition and information analysis are highly automated. However, the decision and action selected by TCAS are "advices" to human pilots, and the pilots may disregard the RA issued by TCAS. Also, TCAS is not given authority for automatic action implementation.

3. HUMAN-CENTERED AUTOMATION

Human-centered automation is an approach to realize work environment in which humans and machines collaborate cooperatively. However, in spite of popularity, there seems to be some ambiguity on what human-centered automation really means: Sheridan (2002) distinguishes ten different meanings of human-centered automation, and argues that contradictions can be found in those "definitions." Among various application domains, it may be aviation for which human-centered automation is defined in the most detailed manner. Aviation has a long history of automation, and has experienced both its benefits and costs (see, e.g., Billings, 1997). The concept of human-centered automation, shown in Table 1, has resulted from studies to resolve the costs of automation (Billings, 1997; ICAO, 1998).

Table 1 Principles of human-centered aviation automation

The human bears the ultimate responsibility for safety of aviation system.
Therefore:

- The human must be in command.
- To command effectively, the human must be involved.
- To be involved, the human must be informed.
- Functions must be automated only if there is a good reason for doing so.
- The human must be able to monitor the automated system.
- Automated systems must, therefore, be predictable.
- Automated systems must be able to monitor the human operator.
- Each element of the system must have knowledge of the others' intent.
- Automation must be designed to be simple to learn and operate.

After Billings (1997) and ICAO (1998).

It is noted that human-centered automation can be domain-dependent: e.g., "human-centered automation for automobile" can be quite different from "human-centered automation for aviation system," and either of these can also be quite different from "human-centered automation for marine vessels" (Inagaki, 2006). Quality of human operators and time-criticality characterize domain-dependence of human-centered automation.

4. ACTIONS IN A GIVEN SITUATION

In Table 1, it is claimed that "Automated systems must be able to monitor the human operator." What the automation may do when it detected the human's inappropriate behavior or performance while monitoring the human? Is it allowed only to give some warnings? Or, is it allowed to act autonomously to resolve the detected problem?

A human's control action or the human's directive to the machine may be classified into three categories: (1) An action that needs to be done in a given situation, (2) an action that is allowable in the situation and thus it may either be done or not done, and (3) an action that is inappropriate and thus must not be done in the situation.

Assuming some sensing technology (or machine intelligence, provided by a computer), two states may be distinguished for each control action: (a) "Detected," in which the computer judges that the human is performing the control action, and (b) "undetected," in which the control action is not detected by the computer.

Figure 1 depicts all possible combinations of a control action and its state. Region A shows the case in which the computer judges that the human operator is (too) late in performing or ordering a control action that must be done in the situation. Region B indicates the case in which the computer determines that the human operator misunderstands a given situation and the control action that he/she takes or requests does not fit the situation. For simplicity of argument, it is assumed in this paper that the computer always knows what action is appropriate beyond just detecting (or not) whether an action is taken.

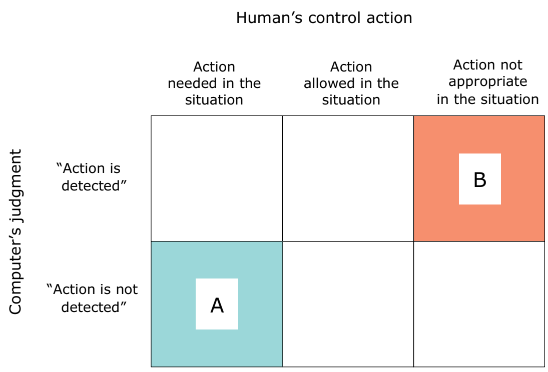

Figure 1 Control action in a given situation

A question that must be asked for the former case (Region A) is whether the computer may be allowed to initiate the control action that the human should have done. A question asked for the latter case (Region B) is whether the computer may be allowed to prohibit the control action that the human is trying to do. In either case, a common issue that must be investigated is whether the computer may be given the authority over the human operator when appropriate, by distinguishing the following two types of authority: (i) authority for the computer to decide and act when a human operator is unable to do what is necessary, and (ii) authority for the computer to prevent a human operator from doing what he/she tries to do.

5. SUPPORT BY WARNING OR BY ACTION

It may be said, in terms of levels of automation (LOA), that a human operator is maintained as the final authority over the machine intelligence (or the computer) only when the LOA for decision and control is positioned at level 5 or lower (see, Table 2). The LOA of a machine-initiated trading of authority for control action is positioned at level 6 or higher. In this sense, so-called human-centered automation principle is violated in such a machine-initiated strategy. However, there are situations or contexts in which a machine-initiated authority trading is indispensable for attaining safety of the human-machine system (Inagaki, 2000; Inagaki, 2006), which suggests the need for precise investigations concerning, "To what extent may the computer be given the authority for decision and control, and in what situations or contexts?"

Table 2 Scale of levels of automation for decision and control

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

After Sheridan (1992; 1999) and Inagaki, Moray, & Itoh (1998)

Some such efforts include: (Inagaki, Itoh, & Nagai, 2006; 2007a; 2007b). By taking an advanced driver assistance system for automobile as an example, they tried to answer the question, "What type of support should be given to a driver when it is determined, via some sensing and monitoring technologies, that the driver's action (or lack of action) may not be appropriate to a given traffic condition?" For cases contained in Regions A and B in Figure 1, two types of support were compared: (a) Warning type support in which an auditory warning is given, and (b) action type support in which an autonomous safety control action is executed to avoid an accident. The warning type support is fully compatible with *human-centered automation*, because the driver was always maintained as the final authority over the automation. Most participants in the experiments accepted the warning type support for either case in Region A or B (Inagaki, Itoh, & Nagai, 2006; 2007a; 2007b). However, the warning type support sometimes failed to prevent an accident when the participants did not respect a warning. A participants' typical and 'reasonable' disregard of a correct warning occurred when the warning was for an invisible (or hard-to-see) object. This fact suggests a limitation of a *purely human-centered automation* design in which the human is maintained as the final authority at all times and on every occasion.

A machine-initiated action taken for cases in Region A may be just to implement a control action that a human has failed to perform in a timely manner. For cases of Region B, machine-initiated control actions are classified into two groups; (a) *hard protection*, in which the human may not override the computer's *corrective* control action initiated based on its judgment that "the human's action does not fit the situation," and (b) *soft protection*, in which the human can override the computer's *corrective* control action, even though it, the computer, judges that, "the human's action does not fit the situation."

The action type support with hard protection can fail to be accepted by participants, although it was successful in accident prevention; e.g., see, (Inagaki et al, 2006; 2007b). A most prominent reason for lack of acceptance was due to the hard protection characteristics in cases when there was a conflict of intentions between the human and the computer. The soft protection type action support, on the other hand, sometimes failed to avoid an accident, when participants misinterpreted why protective action had been triggered and for which object. They had to interpret the situation based on limited information collected within a limited time allowance. An issue is how to design a human-machine interface and interaction for cases when something may be invisible for a human, while visible for the machine.

6. CONCLUDING REMARKS

Whether trading of authority must be human-initiated or machine-initiated has been a crucial research issue in adaptive automation (Scerbo, 1996; Inagaki, 2003, 2008), because trading of authority is essential in function allocation between humans and machines in a dynamically changing environment. This paper has argued that, although human may not be in command in case of machine-initiated trading of authority, such autonomous decision and action implementation are sometimes indispensable to assure safety, efficiency, and comfort of transportation vehicles. It should be noted, however, that the author does not claim that machine-initiated trading of authority is always in need and effective in aviation domain. On the contrary, as has been argued earlier, human-machine collaboration must be designed by reflecting characteristics of the domain considered. Even in the aviation domain, a sensible form of collaboration with machines can be different between the case of pilots and that of air traffic controllers. What is in need is a systematic way of thinking and methodology that can investigate and evaluate design of human-machine collaborations in a quantitative manner with appropriate precision.

REFERENCES

Billings, C.E. (1997). *Aviation automation – The search for a human-centered approach*. LEA.

Inagaki, T. (2000). Situation-adaptive autonomy for time-critical takeoff decisions. *International Journal of Modelling and Simulation*, 20(2), 175-180.

ICAO (1998). *Human factors training manual*. Doc 9683-AN/950.

Inagaki, T. (2003). Adaptive automation: Sharing and trading of control. In E. Hollnagel (Ed.) *Handbook of cognitive task design* (pp. 147-169). LEA.

Inagaki, T. (2006). Design of human-machine interactions in light of domain-dependence of human-centered automation. *Cognition, Technology & Work*, 8(3), pp. 161-167.

Inagaki, T. (2008). Smart collaborations between humans and machines based on mutual understanding. *Annual Reviews in Control*, vol. 32, pp. 253-261.

Inagaki, T., Moray, N., & Itoh, M. (1998). Trust self-confidence and authority in human-machine systems. *Proc. IFAC Human-Machine Systems*, pp. 431-436.

Inagaki, T., Itoh, M., & Nagai, Y. (2006). Efficacy and acceptance of driver support under possible mismatches between driver's intent and traffic conditions. *Proc. HFES 50th Annual Meeting*, pp. 280-283.

Inagaki, T., Itoh, M., & Nagai, Y. (2007a). Driver support functions under resource-limited situations. *Proc. HFES 51st Annual Meeting*, pp. 176-180.

Inagaki, T., Itoh, M., & Nagai, Y. (2007b). Support by warning or by action: Which is appropriate under mismatches between driver intent and traffic conditions?. *IEICE Trans. Fundamentals*, E90-A(11), pp. 264-272.

Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation. *Automation and human performance* (pp.37-63). LEA.

Sheridan, T. B. (1992). *Telerobotics, automation, and human supervisory control*. MIT Press.

Sheridan, T.B. (1999). Human supervisory control. *Handbook of systems engineering and management* (pp. 591-628), John Wiley & Sons.

Sheridan, T.B. (2002). *Humans and Automation: System Design and Research Issues*, Human Factors and Ergonomics Society & Wiley.

US Dept of Transportation & FAA (2000). *Introduction to TCAS II version 7.*