Trust, Autonomy, and Authority in Human-Machine Systems: Situation-Adaptive Coordination for Systems Safety

Toshiyuki Inagaki and Makoto Itoh

September 11, 1996

ISE-TR-96-138

This paper will be presented at the Cognitive Systems Engineering in Process Control, November 12-15, 1996, Kyoto.

Trust, Autonomy, and Authority in Human-Machine Systems: Situation-Adaptive Coordination for Systems Safety

T. Inagaki and M. Itoh
Institute of Information Sciences and Electronics
and the Center for TARA
University of Tsukuba, Tsukuba 305 Japan
inagaki@is.tsukuba.ac.jp, mako@aleph.is.tsukuba.ac.jp

Abstract

This paper discusses an issue how authority and responsibility should be shared between human and machine intelligence for attaining systems safety. Via an experimental approach, the validity of 'situation-adaptive autonomy' concept is investigated, where the concept has been proposed originally by Inagaki through mathematical analyses. Some items for further research are given for implementing the situation-adaptive autonomy which can contribute to safety of human-machine systems.

Introduction

Many large-complex plants in our society are semi-autonomous, where computers control the plants according to directives given by human operators. The configuration can be represented usually by a human supervisory control model (Sheridan, 1992). Among tasks of human operators in the supervisory control configuration, monitoring and intervening are not easy tasks, which raise wide variety of issues on relationship between human and automation. Some of the issues are: lack of situation awareness, complacency, distrust of automation, automation-induced surprises and mode confusion, human-out-of-the-loop problem, which may be found, for instance, in (Muir, 1987; Wiener, 1989; Woods, 1989; Norman, 1990; Amalberti, 1992; Lee & Moray, 1992; Parasuraman *et al*, 1993; Sarter & Woods, 1995).

The concept of 'human-centered automation' is expected to play an important role for resolving the above issues (Billings, 1991; Rouse, 1991). In the human-centered automation, it is said that "a human locus of control is required," which means that the human operator has effective authority as well as responsibility (Woods, 1989). How do we interpret the statement? Does the statement mean that "a human locus of control is required *at any time in any case*?"

Machine intelligence is not powerful enough to handle all abnormalities and the human must be in the system, which yields the human supervisory control configuration. It would be quite appropriate to claim that the human is given effective authority. However, do we claim at the same time that the human must bear the final responsibility at any time in any case? As has been seen in various accidents of aircraft, nuclear/chemical and some other plants, the human operators are put into difficult situations once some malfunction occurs in the plant. Inherent complexity of the plant, poorly designed human-interface or support system, and time or social pressure can make the situations more difficult. Even when human error plays some role in an accident, it might be too easy to assume operator's responsibilities of authority.

We must thus investigate carefully how authority and responsibility should be shared between human and machine (automation). Inagaki (1991, 1993, 1995) proposes the concept of 'situation-adaptive autonomy' where the machine shares responsibility in a positive manner when plant safety is a factor. Some analyses, based on mathematical models show that the machine intelligence may be given right of executing safety-related control actions when necessary, even if the human did not give the machine an explicit directive to do so.

This paper investigates the validity of the situation-adaptive autonomy concept via an experimental approach, because mathematical models are not always powerful enough to incorporate all the human factors. Through experiment, this paper tries to clarify research items for implementing the design of situation-adaptive autonomy which contributes to safety of human-machine systems.

Situation-Adaptive Autonomy

Suppose "a human locus of control is required" in the strictest sense. Then some levels of autonomy in Table 1, levels 6 through 10, are not allowable. Via simple mathematical models, Inagaki (1991, 1993, 1995) has shown that autonomy with levels 6 or higher plays a vital role for attaining safety of the plant. More precisely, Inagaki has investigated the situation in which an alarm (which can be false) has been given and the operator is requested to take safety-control action, if necessary, to avoid possible accidents from occurring. The strategy alternatives investigated there are:

Strategy 1 with autonomy of level 4: Upon receiving an alarm, the operator performs an alarm analysis to check whether the alarm is correct or not. If the operator judges that the alarm was correct, then he commands the computer to shut down the plant. If the operator judges that the alarm was false, then he cancels the alarm.

Strategy 2 with autonomy of level 6: Upon receiving an alarm, the operator performs an alarm analysis to order whether the plant should be shut down or the alarm should be cancelled. If the computer fails to receive any directive from the operator, it shuts down the plant.

Table 1: Levels of autonomy (Shridan, 1992)

1. The computer offers no assistance, human must do it all.
2. The computer offers a complete set of action altematives, and
3. narrows the selection down to a few, or
4. suggests one, and
5. executes that suggestion if the human approves, or
6. allows the humari a restricted time to veto before automatic execution, or
7. executes automatically, then necessarily informs human, or
8. informs him after execution only if he asks, or
9. informs him after execution if it, the computer, decides to.
10. The computer decides everything and acts autonomously, ignoring the human.

Strategies 3 and 4 with autonomy of level 7: Immediately upon receiving an alarm, the computer executes the fault-compensation on the plant before the operator initiates an alarm analysis. The rest is the same as Strategy 1or 2, respectively.

The order relation among the strategy alternatives has been investigated by reflecting human factors, such as credibility of alarms (trust or distrust), time allowed for an alarm analysis or decision making, possibility of knowledge-based mistakes which may be induced by poorly designed human-interface or lack of proper knowledge on the plant, and monetary factors, such as loss caused by spurious shut down of the plant, damage caused by delay or failure of shutting down the unsafe plant. The analyses show the need for Strategies 2 through 4 in preventing an accident from occurring. Inagaki has thus proposed the situation-adaptive autonomy the level of which can be set flexibly and dynamically depending on the situation. However the suggestion has been validated there only through mathematical analyses. This paper investigates the validity of the situation-adaptive autonomy via an experimental approach with simulations.

Human Supervisory Control

The Simulated Plant

In the human supervisory control, we cannot be sure that the human or the machine intelligence has a perfect model of a plant to be controlled. As one of such cases, we investigate a plant control model illustrated in Figure 1. The controlled plant consists of three subsystems. Subsystem A is for adjusting the quality of the 'product fluid' for subsystem B which impose the following requirements: (1) The temperature of the product fluid sent to subsystem B must be kept within the range of 50℃ to 70℃, and (2) the flow rate of the fluid to subsystem B must be within the range of 14 to 22, measured in an appropriate dimension. The fluid which fails to satisfy either one of conditions is not regarded as the proper product. Coming out from subsystem B, the used fluid goes back to subsystem A for renewal. While passing through subsystem C, a portion of the flow quantity is disposed of as waste product. The main pump in subsystem A is activated, when necessary, to refill a suitable amount of fresh fluid into the pipeline.

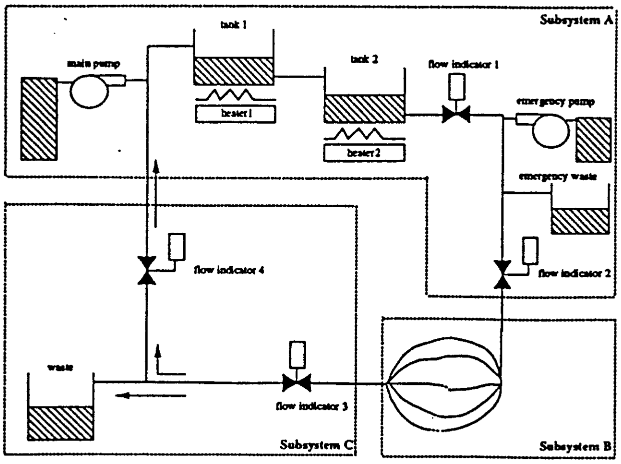

Figure 1: The simulated plant

Malfunctions and Countermeasures

Some malfunctions can occur in the plant.

(1) Pipe rupture. Three levels of pipe rupture can occur: (a) the first stage of rupture, in which 10% of flow quantity is lost, (b) the second stage of rupture, in which the loss of flow quantity becomes large exponentially as time goes on, and (c) the third stage, in which 100% of flow quantity is lost right away. The pipe rupture makes a transition from its first stage to the second in 60 time units, where one time unit corresponds to 1.3 seconds. The second stage lasts 25 time units, and then enters into the third stage. If the pipe rupture of the third stage occurs at a point between tank 2 and subsystem B, the flow rate to subsystem B vanishes within 5 time units if no appropriate countermeasure is taken. We say that an accident occurs at a time point when the flow rate at subsystem B becomes zero.

In the first stage of pipe rupture, the operator can 'repair' it by pressing the 'repair button.' If the rupture is in the second stage, the operator must activate the auxiliary pump to compensate the loss of flow quantity to subsystem B; viz. just pressing the repair button is not enough. Once the pipe rupture enters into the third stage, the operator must shut down the whole plant immediately for avoiding an accident to occur. It is assumed in the simulation that the repair completes five time units after pressing the repair button.

(2) Level indicator failure. Level indicator 2 can give an erroneous reading. The level indicator fails in two ways: (a) The reading is smaller than the true value, and (b) the reading is larger than the true value. In either case, reading error grows linearly as time goes on. Immediately when the

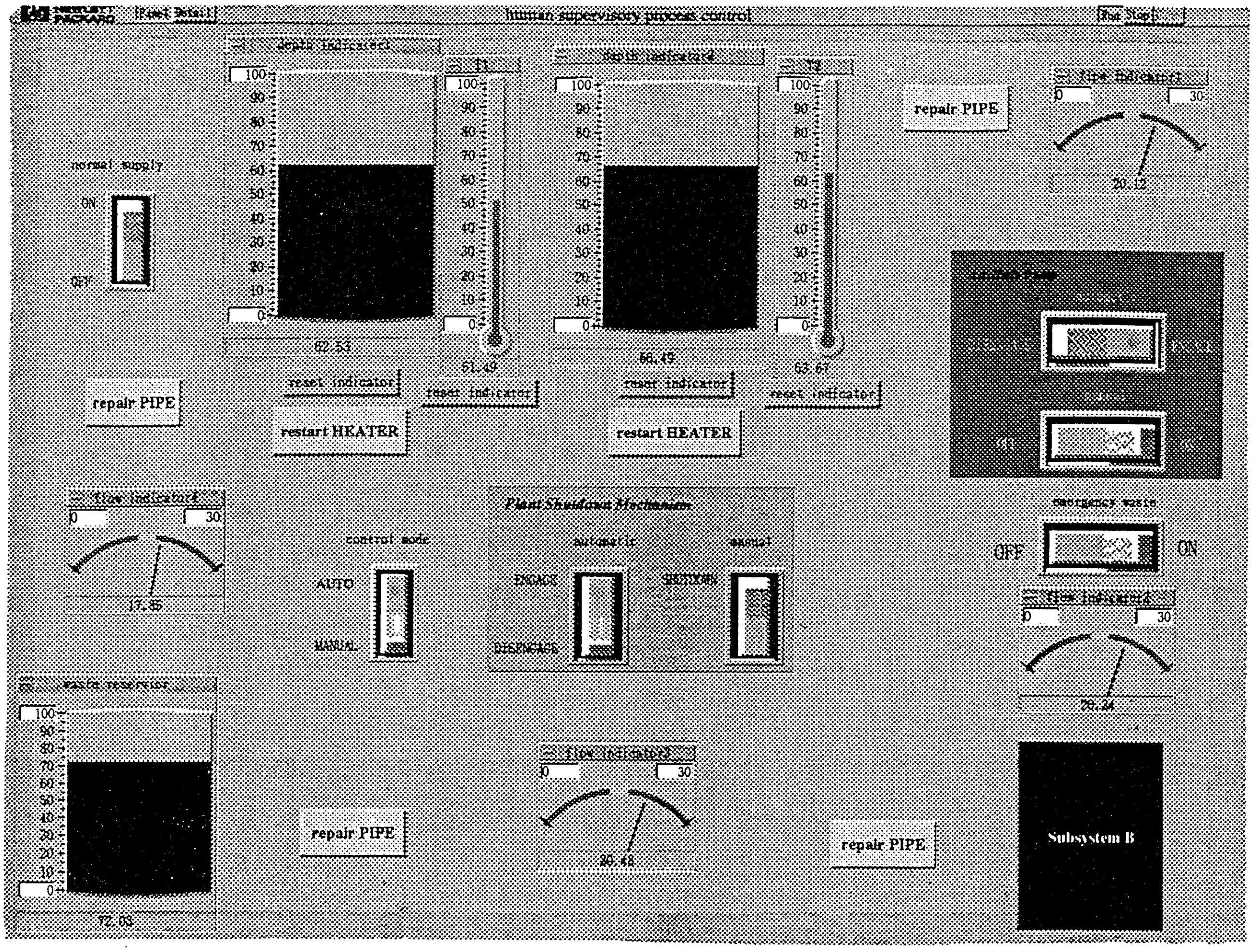

Figure 2: Human-interface of the control panel

operator presses the 'reset button,' the level indicator becomes normal and the reading error vanishes.

(3) Heater failure. The temperature of the fluid at the main pump is around 30℃. It can go up to around 50℃ when heated in tank 1, and further up to about 70℃ when heated in tank 2. Either heater 1 or 2 may lose capability to heat the fluid. The heater comes back to its normal operating condition immediately when the operator presses the 'restart button.'

Human-Interface Design

The operator can guess the plant state only through indicators showing parameter values, such as flow rate, fluid level and temperature. Figure 2 depicts the human-interface of the control panel which is available on the display. The interface has been developed with the Hewlett Packard's VEE for Windows. Some preliminary experiments have been conducted for designing the human-interface.

Tasks Imposed on the Operator

Each operator is requested to perform 'main task' and 'subtask' simultaneously. Main task is to feed as much proper 'product fluid' as possible to subsystem B. To pursue the main task, the operator must: (1) control the main pump appropriately, (2) decide when the auxiliary pump must be activated or stopped, and (3) decide whether the bypass line should be used for discarding fluid in excess.

Two types of sub-tasks are prepared: (1) Transcribing English words or sentences listed on sheets of paper, and (2) solving problems which needs reasoning (see, Figure 3). The former sub-task is considered skill-based, while the latter rule-based or knowledge-based. The sub-tasks are imposed for preventing the operator from concentrating his attention fully on the main task.

Rating of the operator's performance is done based on the performance in main and sub-tasks. Each operator is informed that some monetary bonus will be given if he gets either the highest or the second highest score, which is to make operators 'wise' or 'ambitious,' instead of letting them 'lazy,' according to the wording of (Stassen, Johannsen, and Moray, 1990).

There are seven propositions, A through G. Three students, Rhea, Sarah, and Tiffany have investigated the propositions.

(1) Assuming Proposition D, Rhea proved Proposition C. Based on that, she proved Proposition G.

(2) Aassuming Proposition F, Sarah proved Proposition C. Based on that, she proved Proposition B.

(3) Assuming Proposition A, Tiffany proved Proposition G. Based on that, she proved Proposition D, and she further proved Proposition F.

The students have confirmed that every proof given in (1) through (3) is correct. They are now investigating propositions which are equivalent to Proposition C. How many propositions can they find, excluding Proposition C itself?

Figure 3: An example of problem-solving sub-task

Difficulties in Manual Control

When no fault exists in the plant, manual control of the main pump is not difficult for the operator. Once a malfunction occurs somewhere in the plant, the control task becomes difficult. While the pipe rupture remains in the first stage, the operator can cope with the situation just by controlling the main pump appropriately. If the pipe rupture enters into the second stage, the operator must activate the auxiliary pump, because the main pump is not powerful enough to keep the flow rate at subsystem B above the required LB (lower bound) level: Without the emergency pump, the flow rate at subsystem B falls below the LB level in a few seconds after the pipe rupture entering into its second stage.

The manual control of the emergency pump is inherently difficult, because it is done under abnormal operating conditions of the plant. The point of successful control of the auxiliary pump lies in which flow indicator attention should be allocated to for getting information. The pipe rupture may be detected by reading flow indicator 1, while flow indicator 2 is useful for checking operating conditions of the auxiliary pump. After completion of pipe repair, flow indicator 1 should be consulted again for setting the flow rate to subsystem B at an appropriate level.

If the main and the emergency pump is kept on for a long time period, flow rate to subsystem B may exceed the maximum allowable UB (upper bound) level. Then the operator must open the bypass line to lead some portion of flow quantity to the emergency waste, which means that the pump feeds fluid in vain just to lead it to the waste.

The difficulty of the manual control of the auxiliary pump or the bypass function stems from the time delay. The effect of an operator's control becomes visible on flow indicators about 5 time units after the actual the control action.

Automatic Control Systems

The operator can utilize, if he wishes, any of the following automatic systems for: (1) controlling the main pump, (2) controlling the auxiliary pump, and (3) shutting down the whole plant for preventing an accident from occurring. The activation and deactivation thresholds for the control systems are depicted in Figure 4. The thresholds are set intentionally by the authors to make the automatic control systems not so stupid, but not so wise. It is essential to create automatic control systems which have 'comparable' capabilities to the human operator. If the operator has good skill of control, he may defeat automatic control systems which act according to simple control strategies. If the operator has poor skill of control, he is easily defeated by the automatic control systems.

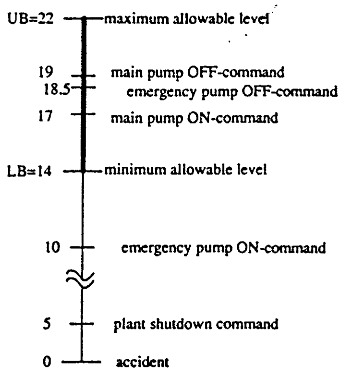

Figure 4: Activation/deactivation thresholds

The operator has a complete right to determine whether to engage automatic control system(s), and which system to use at which situation. The operator can thus set the 'level of autonomy' of the control systems completely freely, where the levels of autonomy are listed in Table 1. It is assumed that no automatic control system fails, which is not informed to the subject.

Results and Observations

Seven students (graduate and undergraduate) participated in the experiment which lasts five days for each subject. Every subject is requested to perform 25 trials in total, where one trial lasts about 6.5 minutes. Every subject is exposed to various scenarios which differ from trial to trial. Prepares scenarios are divided into four categories: (1) no malfunction occurs at any component, (2) one failure occurs at a component, (3) two different components fail independently, where the time between two failures is greater than 70 seconds, and (4) two or more different components fail independently, where the time between two consecutive failures is not greater than 70 seconds. No subject is informed either the categorization of scenarios, or when and how many malfunctions may occur there.

Experimental schedule for five days are shown in Table 2, where sub-task category is also sown. On the first day each subject is given opportunity to: (1) acquire skill for controlling the plant manually, (2) know what happens if a malfunction occurs, and (3) learn capabilities of automatic control systems. The first trial on each day (during Day 2 to

Table 2: Experimental schedule

Day 1: Getting Acquainted
1: no malfunction
# 2: pipe rupture (60)
3: heater 1 failure (90)
4: pipe rupture (60)
5: pipe rupture (60)
under automatic control
Day 2: Transcribing sub-task
1: pipe rupture (150)
# 2: no malfunction
# 3: heater 1 failure (100)
# 4: no malfunction
5: pipe rupture (40)
Day 3: Problem solving sub-task
1: heater 1 failure (250)
2: pipe rupture (130)
3: no malfunction
# 4: no malfunction
# 5: heater 2 failure (180)
# 6: level indicator failure (100) k=+0.2
Day 4: Transcribing sub-task
# 1: heater 1 failure (120)
level indicator failure (200) k=+0.2
# 2: heater 2 failure (120)
pipe rupture (180)
# 3: level indicator 2 failure (70) k=-0.2
pipe rupture (100)
# 4: no malfunction
Day 5: Problem solving sub-task
# 1: pipe rupture: stage 2 (80)
level indicator failure (150) k=+0.2
2: heater 1 failure (110)
pipe rupture (180)
# 3: level indicator 2 failure (70) k=-0.1
pipe rupture (110)
# 4: no malfunction
# 5: level indicator 2 failure (130) k=-0.2
pipe rupture (150)
heater 2 failure (180)
(Note) (i) 'event (y)' denotes the event occurs at time y
(ii) 'k' denotes a coefficient for linear growth of error in the reading of the level indicator

Day 5) is for maintaining or improving manual control skills of the subject.

Table 3 shows scores and ranking of subjects. The experiment has raised issues for further study to implement well-coordinated human-centered automation with situation-adaptive autonomy. The issues are described in the following.

Situation Awareness

One of the interesting observations is that the type of sub-tasks has little influence on the time for the operators to detect pipe rupture (see, Table 4), while it seems to have some effect in case of level indicator failure (see, Table 5). Subjects usually pay attention more to flow rate than to fluid level, because the former gives primary information for controlling pumps. Sub-tasks of transcribing words or sentences are of skill-based (see, left-halves of Tables 4 and 5), but were likely to make subjects' eyes away from plant information given on the CRT, while subjects could take a look at the CRT occasionally even while they were solving problems (see, right-halves of Tables 4 and 5). That may be a possible explanation for influence of sub-task type on the time to detect level indicator failure.

Table 3: Subjects' scores and ranking (Ψ: accident, '*': accident was prevented by the automatic control system)

	ranking	score									
		trial 2		trial 3		trial 4		trial 5		trial 6	
Day 2 (subtasks of transcribing words)	1	MA	5466	TO	5071	MI	9096	MA	9301		
	2	SH	5460	IN	4869	TO	6680	TO	7368		
	3	MI	5367	MA	4568	MA	5889	IN	7139		
	4	KA	4949	SH	4480	IN	5454	KA	5430		
	5	IN	4948	MI	4350	KA	4980	SH	5121		
	6	TO	4771	KA	4248	TS	4954	TS	4389		
	7	TS	4722	TS	2957	SH	3498	MI*	1795		
	average		5098		4363		5793		5792		
Day 3 (subtasks of problem solving)	1	MA	9209	TO	9411	MA	9386	TO	9262	TO	9409
	2	MI	9052	MI	7045	MI	9382	MA	9224	IN	9408
	3	SH	8203	KA	6938	SH	9257	MI	9182	MI	9408
	4	IN	8040	IN	6816	KA	8939	IN	8941	MA	9378
	5	TO	6413	SH	6761	TS	8938	KA	8544	SH	9134
	6	TS †	0	MA	5407	IN	8626	TS	8520	KA	8940
	7	KA †	0	TS	4950	TO	7403	SH	8519	TS	6958
	average		5845		6761		8847		8885		8948
Day 4 (subtasks of transcribing words)	1	IN	8384	TO	9338	TO	9406				
	2	SH	8224	MA	5378	IN	8950				
	3	TO	7219	IN*	4398	MA	5407				
	4	MA	5043	MI	3964	MI	5265				
	5	TS	4157	SH	3861	TS	5013				
	6	MI*	3535	KA	3570	KA	5004				
	7	KA*	2527	TS*	2387	SH	4951				
	average		5584		4699		6285				
Day 5 (subtasks of problem solving)	1	TO	7037	MA	7332	TO	9403	MA	9134		
	2	TS	6599	TO	7283	SH	8952	TS	8845		
	3	MI*	6260	IN	7100	MA	7399	IN	8231		
	4	MA	5217	SH	5564	TS	7395	SH	7625		
	5	IN	5045	KA	4601	MI	7370	TO	7248		
	6	SH	3978	TS	4514	IN	7160	KA	6265		
	7	KA	3746	MI †	0	KA	6941	MI	4789		
	average		5412		5199		7803		7448		

Consecutive failures of level indicator and pipe can prolong the time to detect malfunctions (see, Figure 5). A typical story happened on an 'ordinary' operator was: "Level indicator 2 failed at some time point in a mode to give reading smaller than a real value. The magnitude of reading error grows gradually with time, but the operator did not recognize that. In the meantime pipe rupture broke out, and flow indicator 1 gave reading smaller than usual. Level indicator 2 and flow indicator 1 happened to coincide in giving low readings. Actually, the subject thought that the coincidence proved shortage of fluid in tank 2. He should have taken seriously the phenomenon that the level of tank 1 was increasing at that time; the main pump was turned on by the automatic control system which had detected the low flow rate. The subject noticed the pipe rupture at last after it entered into the second stage where loss of flow quantity

Table 4: Time elapsed before pipe rupture was repaired, where imposed sub-task was: (1) transcribing words (Day 4 Trial 2), or (2) problem-solving (Day 5 Trial 2)

	time elapsed before pipe repair	
	transcribing sub-task Day 4, Trial 2	problem solving sub-task Day 5, Trial 2
MA	5	6
TO	14	23
IN	29	5
SH	46	64
TS	65	66
KA	72*	74*
MI	73*	69*
average	43.4	43.8

Table 5: Time elapsed before level indicator was repaired, where imposed sub-task was: (1) transcribing words (Day 4 Trial 1), or (2) problem-solving (Day 5 Trial 1)

	time elapsed before level indicator repair	
	transcribing sub-task Day 4, Trial 1	problem solving sub-task Day 5, Trial 1
MI	23	29
KA	24	29
TO	30	14
MA	43	12
TS	48	52
SH	68	33
IN	90	35
average	46.6	29.1

became larger. He did not recognize the level indicator failure for about three more minutes after detecting the pipe rupture."

A 'smart' subject could find pipe rupture and level indicator failure almost immediately in the above situation. His success story was: "When he found the coincident decrease in readings on the flow indicator 1 and level indicator 2, he compared the reading of flow indicator 1 with those at downstream flow indicators 2 and 3. Due to time delay, the effect of the pipe rupture had not reached flow indicators 2 and 3 yet at that time. The subject thought that pipe rupture was occurring at somewhere upstream of flow indicator 1, and pressed the pipe repair button. Moreover he found the failure of the level indicator 2 based on his knowledge on the relationship between fluid level in tank and flow rate. He thought that reading of level indicator 2 was 'too low' judging from the reading of flow indicator 2." We would say that the subject had a very good mental model on the plant dynamics.

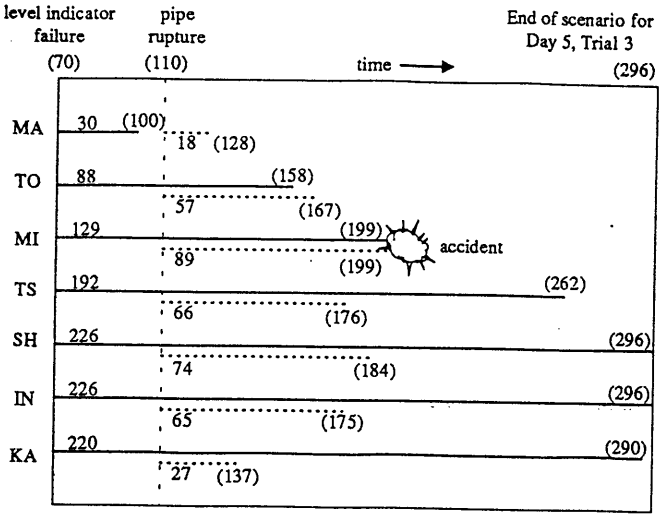

Figure 5: Time before mulfunctions were detected under circumstances with consecutive failures

Automatic Control Systems Hide Abnormalities in the Plant

If the automatic system is engaged for controlling the main pump, it can mask pipe rupture in its early stage, because the automatic system can cope with the situation by making the main pump work to compensate the lost portion of flow quantity. The operator can find abnormality only when he has a good mental model on relationship between the readings in flow indicator and level indicators at tanks: If the flow rate is not big, but if the fluid level is high, then a pipe rupture may be suspected. Large value of reading in level indicators may due to the main pump's continual and automatic supply of fluid into tanks.

In addition to the issue of mental models, it may be a factor "who sets the reference value for controlling the main pump." In case of manual control, the operator sets the reference value by himself, and thus he can detect pipe rupture within its early stage by recognizing 'discrepancy' between his intention and the realities. In case of automatic control, on the other hand, the operator may not know at which point a reference value is set and thus the intention of the automatic system. This relates to the situation awareness, which has been described earlier.

Trust and Dependence on Automation

Subjects who are not very sure of their manual control skills tend to use automatic systems for controlling main and auxiliary pumps. When sub-tasks are imposed on subjects, no subject, including skillful subjects, were successful in avoiding situations where the flow rate at subsystem B falls below the LB level.

Pipe rupture of its second stage requires the auxiliary pump. In the manual control of the auxiliary pump, every subject was late, compared with the automatic control cases, in executing starting up and/or shutting down procedures for the auxiliary pump. That means that the subjects supplied to subsystem B too little fluid in the early second stage of pipe rupture, and too much fluid in later second stage of pipe rupture (viz., beyond the maximum allowable UB level). When the flow rate was about to exceed the UB level, subjects should have made bypass line open to lead excess fluid to the waste, but most of them could not do that under the burden of sub-tasks. If the auxiliary pump was in its automatic control mode, it was rare that the flow rate to subsystem B exceeded the UB level even under pipe rupture of the second stage with sub-task burdens.

More definite usefulness of the automatic control systems was proven in preventing accidents. As are shown in Tables 3 and 4, some subjects have experienced that the whole plant was shut down by the automatic system. The subjects did not or could not shut down the plant by themselves even though the flow rate to subsystem B was approaching to zero. Subjects, who did not engage the automatic system for shutting down the plant in case of emergency, got accidents (see, Table 3 and Figure 5).

The above results tell that automatic system can play a vital role for accident prevention when the human operator fails to take an appropriate countermeasure by himself. As a matter of fact, subjects have exhibited tendency to engage automatic systems more for sure after they experienced accidents.

Distrust of Automation

Several subjects said that they could not fully trust the automation. The following comment was made against the automatic control algorithm for the main pump: "The main pump seems to be reluctant in starting operation even when the flow rate is falling down to the LB, which makes the human anxious. On the other hand, the main pump does not stop even though the flow rate is approaching to the UB. The control strategy taken by the automatic system is too aggressive. This kind of algorithm make us nervous."

Critical comment were made also on the automatic control algorithm for the auxiliary pump: "Why not let the auxiliary pump start operation a bit earlier than in the current version of the control algorithm? Suppose we see that the reading of flow indicator 2 is decreasing, which might be due to pipe rupture of the second stage. While operating the plant in the automatic control mode, we sometimes cannot be sure that the auxiliary pump will definitely start its operation." It should be noted here, however, that changing condition for making the pump active earlier than the current version does not resolve the problem completely. Distrust of automatic control for the auxiliary pump stems from the existence of time delay before the control effect becomes visible after its execution.

Unintended Use of Automation

One of subjects used the automation in a way which has never been imagined by the designers. He kept the power button for the main pump at the ON position, and controlled the pump 'manually' by using the switch for engaging/disengaging the automatic control system. The subject utilized the characteristic property that manual control switch will be disabled while automatic control system is engaged, which can be a cause of 'mode confusion.'

We thought, before conducting experiments, that subjects would use automatic control systems more while they are doing sub-tasks of problem solving than while they are doing simple and skill-based sub-tasks of transcribing words or sentences. However the opposite was the real case in the simulation study. Transcribing sub-tasks have prevented the operator from monitoring plant, but problem solving sub-tasks has not, which suggests that abstraction-level of sub-tasks may not necessarily be a critical factor in evaluating the degree of mental workload imposed on the operator.

Mode Confusion and Description Error

One of subjects detected pipe rupture when it entered into the second stage, and he pressed the button to start up the auxiliary pump. Several seconds later the subject noticed that no fluid was fed by the auxiliary pump, and pressed the button again and again to start up the pump. The subject did not notice then that he had engaged the automatic system for controlling the auxiliary pump and that the ON-OFF button for manual control had been disabled. Four seconds passed before the subject could finally recognize which control mode he was in. The similar mode confusions were observed in the control of the main pump.

With an intention to engage the automatic system for shutting down the whole plant in case of emergency, one of subjects pressed the button designed for engaging the automatic control system of the main pump. The subject knew very well the locations of the two different buttons. The subject was under mental pressure caused by sub-tasks. Some other subject said that he could not remember which button he had pressed, even though he remembered he did press some button.

Concluding Remarks

The simulation study has proven that the situation-adaptive autonomy plays a vital role for assuring system safety. It is neither wise to assume that autonomy must be fixed a single level, nor that computer should be an 'always obedient subordinate' which is allowed to do what is ordered only when it is ordered. System safety may not be attained if we interpret the concept of the human-centered automation as "human locus of control is required *at any time in any situation*." As has been seen in the experiment, situations can occur in which operator's detection of abnormalities or execution of countermeasure happens to be late.

The situation-adaptive autonomy, however, is still in a premature phase. The experiment raises research issues for implementing a trustworthy mechanism for the situation-adaptive autonomy. Among them the authors are now investigating on the following topics, where some results can be found in (Itoh and Inagaki, 1996):

(1) Developing human-interface for aiding situation-awareness, which can provide:

(a) predictive information for coping with time-delay before effects of control actions become visible,
(b) compactly integrated and clearly explaining information, such as one implemented by an ecological interface approach (Vicente & Rasmussen, 1992), and
(c) information to let the human know operating conditions and intention of automated systems.

(2) Developing mechanisms for resolving distrust of the automated systems; viz., mechanisms for:

(a) preventing 'automation-induced surprises' (Sarter & Woods, 1992, 1994; Wickens, 1994) from occurring,
(b) allowing the operator to adjust automatic control strategy so that it fits to him,
(c) providing 'intelligent' alarm messages where credibility of alarms is visualized, and
(d) flexible coordination of humans and machines by taking dynamic behavior of trust into consideration.

Acknowledgments

This work has been partially supported by the Center for TARA (Tsukuba Advanced Research Alliance) at the University of Tsukuba, and Grant-in-Aid for Scientific Research 07650454 and 08650458.

References

Amalberti, R. (1992). Safety in process-control: An operator-centered point of view. *Reliability Engineering and System Safety*, 38, 99-108.

Billings, C.E. (1991). *Human-Centered Aircraft Automation: A Concept and Guidelines.* NASA TM-103885.

Inagaki, T. & G. Johannsen (1991). Human-computer interaction and cooperation for supervisory control of large-complex systems. In Pichler, Moreno Diaz (Eds.) *Computer Aided System Theory*, LNCS 585, Springer-Verlag, 281-294.

Inagaki, T. (1993). Situation-adaptive degree of automation for system safety. *Proc. 2nd IEEE Int. Workshop on Robot and Human Communication*, 231-236.

Inagaki, T.(1995). Situation-adaptive responsibility allocation for human-centered automation. *Trans. SICE of Japan*, 31(3), 292-298.

Itoh, M. & T. Inagaki (1996). Human-interface design for situation awareness, *Proc. 5th IEEE Int. Workshop on Robot and Human Communication* (to appear).

Lee, J. & N. Moray (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10), 1243-1270.

Muir, B.N. (1987). Trust between humans and machines, and the design of decision aids. *Int. J. Man-Machine Studies*, 27, 527-539.

Norman, D.A. (1990). The 'problem' with automation: inappropriate feedback and interaction, not 'over-automation.' *Phil. Trans. R. Soc. Lond*, B327, 585-593.

Parasuraman, R., R. Molloy, & I.L. Singh (1993). Performance consequences of automation-induced "complacency." *Int. J. Aviation Psychology*, 3(1), 1-23.

Rouse, W.B. (1991). *Design for Success.* Wiley.

Sarter, N. & D.D. Woods (1992). Pilot interaction with cockpit automation. *Int. J. Aviation Psychology*, 2(4), 303-321.

Sarter, N. & D.D. Woods(1994). Pilot interaction with cockpit automation II. *Int. J. Aviation Psychology*, 4(1), 1-28.

Sarter, N. & D.D. Woods (1995). Autonomy, anthority, and observability: Properties of advanced automation and their impact on human-machine coordination. *Proc. IFAC MMS*, 149-152.

Sheridan, T. (1992). *Telerobotics, Automation, and Human Supervisory Control.* MIT Press.

Stassen, H., G. Johannsen, & N. Moray (1990). Internal representation, internal model, human performance model and mental workload. *Automatica*, 26, 811-820.

Vicente, K.J. & J. Rasmussen (1992). Ecological interface design: Theoretical foundation. *IEEE Trans. SMC*, 22, 589-606.

Wickens, C.D. (1994). Designing for situation awareness and trust in automation. *Proc. IFAC Integrated Systems Engineering*, 77-82.

Wiener, E.L. (1989). *Human Factors of Advanced Technology ('glass cockpit') Transport Aircraft*, NASA TR-177528.

Woods, D. (1989). The effects of automation on human's role: Experience from non-aviation industries. In Norman and Orlady (eds.), *Flight Deck Automation: Promises and Realities.* NASA CP-10036, 61-85.