HUMAN-AUTOMATION COORDINATION MEDIATED BY RECIPROCAL SOCIALITY

Tetsuo Sawaragi

Graduate School of Engineering, Kyoto University
Yoshida Honmachi, Sakyo, Kyoto 606-8501, Japan.

Abstract: Human and computer subsystems should be structured and designed to work in mutually cooperating ways guaranteeing a user's usability. For this purpose, progressive system redesigns are needed with respect to human computer interactions to increase system reliability and transparency by increasing human-system interactions and especially a human user's proactive participation, rather than by eliminating the human out of the loop. Such a *socially-centered view* on the human-machine system design regards a human and an automated agent as *equivalent* partners, and through their mixed-initiative interactions some novel relations of mutual dependency and reciprocity would emerge as well as flexible changes of role-taking are expected. After surveying the problems incurred by the conventional technology-centered automation in a variety of fields, we put an emphasis on the fact that a concept of *sociality* is really needed to form the ideal relations of human-automation and to let them emerge out of intimate interactions. *Copyright © 2002 IFAC*

Keywords: Interface agent, usability, human-system interactions, human-centered design, socially-centered automation.

1. INTRODUCTION

In contrast to the conventional technology-centered design of artifacts (*e.g.*, automation systems), now we should consider ideal human-computer relationships with respect to their "interactions". This means that the human and computer subsystems should be structured and designed to work in mutually cooperating ways guaranteeing a user's usability. For this purpose, progressive system redesigns are needed with respect to human computer interactions architecting to increase system reliability and transparency by *increasing human-system interactions* and especially a *human user's proactive participation*, rather than by eliminating the human out of the loop. Such a *socially-centered* view on the human-machine system design regards a human and an automated agent as equivalent partners, and through their *mixed-initiative* interactions some novel relations of *mutual dependency* and *reciprocity* would emerge as well as flexible changes of role-taking are expected.

In the design of human-automation systems, cognitive behaviors and strong affective elements come into play through the human user/operator and serious deficiencies can become apparent after system is delivered and is put to work. Wherein, extensive user participation in design stages is a key for success in both innovation promotion and integration of human factors, which leads to "mutual design and implementation" to facilitate user acceptance of computer systems.

Toward the final goal of establishing "mutual knowledge of intent" between a human user and automation, currently formal arrangements for incorporating human factors both in system design and system development are sought for, and as a means for that, exploitation of the computer as an aid in system design conjunction with user participation is truly needed. In a word, automation systems should not be designed only from the perspectives of "utility" attained in the final products isolated from the user's participation. Rather, they should be designed from the perspectives of "relations" and "processes" emerging between the artifacts of automation and the human user. In order to keep the users in their active and continuous commitment loops, we have to design the emergence of novel

views and perspectives out of the increased and varied interactions between a human user and artifacts so that a user's continuous participation and commitments can be maintained.

In this article, after surveying the problems incurred by the conventional technology-centered automation in a variety of fields (Sawaragi *et al*. 1999), we put an emphasis on the fact that a concept of sociality is really needed to form the ideal relations of human-automation out of intimate interactions. Then, we present an *interface agent* as a semi-automated artifact enabling that, and discuss about its social relationships with a human user.

2. DOUBTS ON HUMAN-CENTERED AUTOMATION AND ITS ALTERNATIVE: SOCIALLY-CENTERED AUTOMATION

In the conventional design of human-machine systems, the designer of the system has taken an *external* position lying outside of the interactions made between the human user and the automated system. He has been an *external* observer and his task has been on how to design the *optimized* interaction between them based on the technology-centered idea. In the past, a human-centered design concept was proposed (Billings, 1997). However, this is still a mirage and its practical realization is thought to be suspicious (Sheridan, 1997). As far as a human-centered design concerns, this is no more than a conventional design principle; how to optimize the predicted interactions between the human and the machine regarding a human user as a *passive* information processing agent. However, interactive domain formed by the human and the automated system as well as by the humans mediated by the automated system is too complex to be predicted exhaustively and to be optimized at the design stage. Moreover, proactiveness and competency are characteristics specific to any human users, and those may affect the ways of interacting with the others. Wherein, a number of novel and unknown relations do emerge among them, some of which may indeed lead to the bottom-up design of ideal coordinations, but some others bring about a catastrophe (*i.e.*, a disorder). Depending upon how they are used by the human, the artifacts do not always function as the designer originally expects, rather ironically they may contribute to enlarging the discrepancy between the human and the machine.

The conventional design principle of human-machine systems has been based on the conception that a human user is a likely source of significant variance in system performance, then it is better to control this source of variance earlier and to minimize such human factors out of system development. However, a human itself is too complex to be controlled, and the human is at the same time a source of *creativity* that sometimes exceeds out of the designer's restricted views. Thus, the system should be designed so that it can let a

human and an automated agent share knowledge about the other's operations and functioning, intent, and plans.

It is widely accepted that extensive and proactive user participation is a key for success in both innovation promotion and the integration of human factors. Moreover, as a true partner of the user an automated system also has to have an analogous capability of "sociality" to be embedded within the interactions with the human user and to form a creative interactive domain together with him.

Anyway, human and computer should be looked as two parties coming from the same positions and having different expertise. The human and the automation system must be regarded as *equivalent* partners, rather than one of them is regarded as superior to the other. It may be still usual that there happens a contradictory situation where the designer and the user of the automation system meet conflicts, but this should be allowed, if it were not critical. What is more important here is how to let them cooperate with each other to change a current conflict status into a new coordinated one that can lead to a safer system status and the better operations. In this sense, we have to extend our views on a human-computer relationship toward a socially-interactive relationship between the human and the computer to establish a *consensual* domain between the automated agents and the human.

3. EMERGING PROBLEMS OF HUMAN-AUTOMATION COORDINATION

A. Robotics Domain

In the current market, there appeared a distinguished shift of preferences of the consumers seeking for the *social robots*. Typical examples of that are "pet robots" and "cohabitant breeding game software", both of which are based upon the interactions between the products and the human user and the users seek for enjoying the conversation, both in verbal and nonverbal communication (Sawaragi *et al*., 2001b). The common properties embedded in those products are:

- A lack of technology is supplemented by the user's proactive engagement.
- Users can actually feel that they are committing with the artifacts.
- Entrainment of the human user within the interactions with the artifacts is designed.
- Design of human-intervening "events", rather than the "products"
- Targets of the consumers are people who seek for experiencing, rather than the people whose needs and goals are explicit and preexisting definitely.

In a word, those products are oriented towards the design of "relations" and/or "process" between the artifacts and the human user, rather than the products isolated from the user's participation. In order to keep the users in their active and continuous commitments, we have to design the emergence of novel views and

perspectives out of the increased interactions between the human user and the artifacts, and the prerequisite therein is the user's continuous *participation*; becoming engaged, and being embedded in a situation.

This is a typical style of a *social* interaction, and is typically seen in *turn-taking* between a mother and her baby in humans. Such an interaction emerges without any mechanisms, which are explicitly controlling turn-taking between them: a mother responds to her baby's pauses in sucking with jiggling in order to encourage the infant to resume sucking. The success of this emergent turn-taking (jiggling, sucking) relies fully on the mother's *interpretation* of the baby's behavior, although the mother received a message that was never sent.

Learning for the social robot, *social learning*, is to improve such appropriate selections based on reinforcement that is brought about by the individual perception of progress relative to the *current* goal (not the absolute one), and/or by evaluating performance according to its *own* perspective based on the use of its resources. For a behavior-based system, resources are the behaviors available for controlling the actions of the robot. Assuming that behaviors are programmed with an initial policy, the objective is to try to find some regularity in the interactions between the robot and the human reacted in the robot's use of behaviors over time.

With respect to the meanings that the social robot gets to have, theoretical model is needed to explain how an autonomous agent may create and dicover new meanings. The agent is autonomous in the sense that its ontology is not explicitly put in by a designer, nor is there any explicit instruction. Meaning is defined as a conceptualization or categorization of reality, which is relevant from the viewpoint of the agent. Meanings need not be expressed through language, but may take many forms depending on the context and nature of the situation concerned.

B. Aviation Domain

Typical breakdowns of human-automation coordination are seen in aviation automation. Actually, there occur human errors induced by the difficulties in interactions between flight crews and cockpit automation. FAA's Human Factors team pointed out some accidents specific to such a new type of airplanes in their report in 1996. The most important point these accidents imply is that the common cause of those accidents is a "failure of situation awareness". Pilots could not be aware of the situation that the airplane encounters due to complex automated systems and lack of their feedback, which led to CFIT (Controlled Flight Into Terrain) accidents. From another viewpoint, this kind of accident comes from the conflicts between designers and users of the automated systems. The designers want to design airplanes in the manner of "technical-centered" automation, and the users hope aircraft to be designed according to "human-centered" automation.

In normal operating modes, such a conflict may be hidden and not noticed, but in an emergency or a time-critical situation, this conflict emerges and leads to a catastrophe.

The impact of automation on the change of communication styles among the humans mediated by kinds of automation tools is also a serious problem. Especially in the field of aviation, introduction of automated systems to the aircraft and air traffic control may have effects on the styles of coordination between the controllers and the pilots. Replacement of the conventional "party-line" communication with the "peer-to-peer" communication would seem to improve efficiency of communication, but at the same time, it has a risk of losing some "invisible relations" that are not consciously noticed both by the designers nor the users so far, and may prevent the two parties from sharing their situation awareness.

Second, concerning with the involvement of the human pilot in the control loop, pilots would like to stay in the loop. We often discriminate concepts of "hard" protection and "soft" protection; along the former concept the machine system should override the human any time, while the latter insists that the machine's task must be restricted to presenting warnings and suggestions, and the final decisions should be left to the human. The pilot's preference to either of soft or hard protection may depend on the degrees of trust to his own skills and other factors including the quality of the training, the geographic conditions of the airports, etc., but anyway the main reason why they want to stay in the loop is that they want to keep their active mindedness during the flight, and this is the most important factor to keep the pilot's high vigilance and good situation awareness as well as to prevent their skills from degrading. Actually during the flight the cross-checking among the crews is made and this is very effective to keep their vigilance.

C. Design and Manufacturing Domain

In a field of manufacturing, an automation tool has contributed much with respect to a design phase. Most representative one is a CAD (Computer Aided Design) tool automating drawing tasks and parametric design tasks. The next generation CAD will come to contain product data in addition to drawing data so that they could be shared in real-time among distributed design group members as well as among different sections (*i.e.*, designers, production engineers, service engineers dealing with the maintenance, etc.).

Advanced CAD tools should have a capability of supporting a designer's creative task in a coherent way with the human expert designer does, and it has to enable not only design data but also design knowledge to be transferred among the design participants. Wherein, designers must be able to understand the intention of the original design precedents to successfully use them, but current design automation tools are insufficient for

supporting such a task since they do not contain any design intention, decisional process information nor design knowledge. Design modification with lack of such expertise and with mis-understanding of the design precedents would lead to the design causing a malfunction of a product, which incurs a lot of time and expense cost.

The typical catastrophe caused by this was the JCO criticality accident happened in 1999, Japan. This happens because a facility originally assigned to the task was intentionally replaced with other available ones by the worker at the site. This was to increase the efficiency of the work. Those alternative facilities were indeed functionally equivalent, but the safety design constraints implied in the original ones were not understood by the workers (*i.e.,* users), which leaded to the accident.

The generic lessons learned from the JCO criticality accident are as follows:

[1] Do not assume workers simply follow procedures, even if the tasks seem well routinized. Any human being would have some kind of model of the system.

[2] In some situations, workers are forced to create a new system and procedure to satisfy requirements, which is to be based on their mental model.

[3] Support workers to form an appropriate mental model and to perceive the system state through making visible the system state in terms of the safety boundaries as well as the system functional structure.

That is, we have to note the characteristics of the human-in-the-loop system with respect to the emergence expected to occur therein. Especially, how to let the workers discover the safety boundaries by themselves is of critical importance, though the current technology-centered automation has paid little attention with respect to this. We believe this is would be only possible by allowing the human workers a variability of their actions and by letting them share deep design knowledge of the designers.

In Japan such deep knowledge of design expertise have been transferred from expert designers to novice designers on the job training (OJT) directly through a man-to-man communication. However, current severe economic recession causes frequent re-assignment of expert designers to different work sections, which makes the transfer of knowledge quite difficult. We are now seeking for alternative styles of knowledge transfer such as schooling education and formal documentation of design expertise, both of which are not going well as expected. The main reason why the experts are unwilling to transfer their knowledge is very simple; it is impossible to teach or formally describe design expertise by isolating that from the practice of design activity. In order to offer an effective support to a new design problem, design precedents must be stored together with implicit problem solving know-how used and realized in them.

The problem of knowledge transfer may be caused by the frequent re-assignments of the experts designers. But on the other hand, fluid participation in a variety of tasks or regular turnover of the team members from experts to novices do contribute much to enriching experts' adaptability and to keep the level of organizational knowledge that is learned by the organization itself rather than within the individual members. The most big problem causing the transfer of expertise is individual designer's "closed" attitude keeping him/her from learning new things and/or new ways of thinking (*i.e.*, adherence to what they are getting accustomed). How to get rid of such barriers and make individuals open to new ideas will be critical.

D. Process Monitoring Automation Domain

An ideal future human-automation relationship must be the one like a negotiation. That is, human and automation should be in "win-win relationship" symbolizing a fact that the human and the automation system are the equivalent partners, rather than a supervisor nor a supervisee. It is usual that there happens a contradictory situation where the designer and the user of the automation system meet conflicts. If they keep insisting on what they believe and do not change their views, it may lead to a catastrophe. To avoid that they have to cooperate with each other to change a current conflictive situation into a new coordinated one that can lead to a safer plant status and the better operations. Human and computer should be looked as two parties coming from the same positions and having different expertise. For instance, computers can know about the feasibility of certain solutions in a physical sense and it can search a huge state space, whereas the human comes to the negotiation table with ideas about what to prefer in terms of criteria. Thus, negotiation should be done between them by exchanging those expertise defined on the feasibility space and the preference space back and forth so that they can reach to the solution.

E. General Discussions for Crossover Domains

In concluding the above, we would like to highlight a number of common problems that are clarified as mentioned so far.

First, the above highlights what the designers of human-machine systems have to think of when a new automation is introduced to a particular task domain; how the communication between the human and the automated system should be designed, and how the automation effects on the conventional communication styles of the humans. And these issues can be translated into the following more general query on "Who is responsible for the design of coordination; a human user, a system designer or an artifact of automated agent ?"
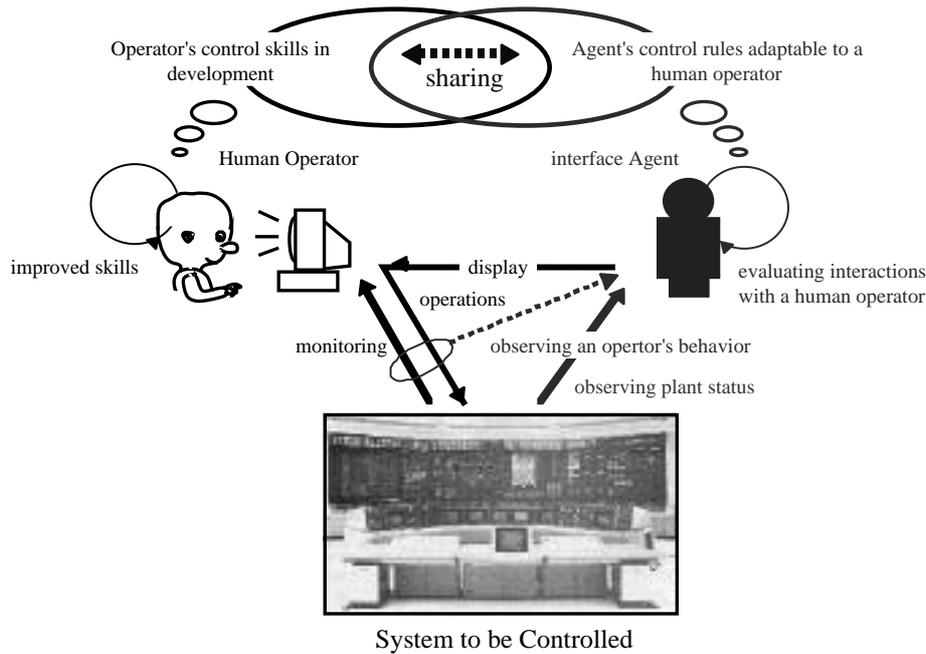
Fig.1 Interface agent as a human associate

Another point commonly depicted by the above is to establish "mutual knowledge of intent" among those three parties (*i.e.*, a user, a designer and an automated system). This is not novel idea *per se*, but what is to be emphasized here is that "mutual knowledge of intent" is not any product to be obtained and shared, but is a continuous process of getting more and more aware of their partners (*i.e.*, an emerging property). Extensive and proactive user participation is a key for success in both innovation promotion and the integration of human factors. Moreover, as a true partner of the user an automated system also has to have an analogous capability of "sociality" to be embedded within the interactions with the human user and to form a creative interactive domain together with him.

4. INTERFACE AGENT AS A HUMAN'S ASSOCIATE

A. Overview of Interface Agents

An interface agent is a semi-intelligent computer program that can learn by continuously "looking over the shoulder" of the user as he/she is performing actions against some complex artifacts and is expected to be capable of providing the users with adaptive aiding as well as of alternating the activities instead of a human (Fig.1). Thus, an interface agent is not a fixed program, but must be *co-evolve* with the human user as it interacts with the human user.

As a human partner, an agent must be so *open* and *social* that the human can grasp what and how the agent is thinking, apart from whether he/she thinks that it is true or not. Instead of taking conventional artificial intelligence approaches for an interface agent, we proposed a new designing method using

the *ecological approach*, since we would like to stress the importance of the agent's ability to create reciprocal relationships with the human user (*i.e.*, a social ability), rather than its knowledgeable ability enclosed within the agent itself.

B. Technical Issues towards Opens and Social Automation

It is well recognized that collaboration requires the establishment of the following three kinds of relations among the participants; *augmentative* (for compensating for mutual incompleteness), *integrative* (of different views), and *debative* (in sharing critics and criticism). Through either of these relations, the agent has to be continuously interactive with the human.

Another fundamental issue to be addressed in a design of interface agents is its *reactivity*; it has to deal with real-time problems that evolve and require solution in real-time, and with multi-tasking systems and missions. The actual application domains having such real-time, multi-tasking characteristics are; en-route air traffic control, helicopter pilot-vehicle interaction, naval command and control, and aerospace, process/manufacturing control, telecommunication, medical applications. Common characteristics therein are that tasks may interrupt one another, and a given task may be interrupted and resumed several times as the person/team copes with the on-going sequence of real-time events in the external problem environment. Thus, agents have to be able to think about how conflicting demands for attention are resolved, when two or more tasks simultaneously demand attention, since the agent as well as its human partner are always performing one and only one cognitive task at any given time (*i.e.*, a resource management task).

C. Our Approach to Interface Agent Design

Instead of taking conventional artificial intelligence approaches for an interface agent, we are adopting a new designing method using the *ecological approach*. So far we have done a number of experiments on collaboration between a human user and an interface agent. These results were discussed from the following perspectives.

- The methodological aspects of ecological approach of Brunswik's Lens Model Analysis for measuring and quantifying sociality between the two decision making agents (*i.e.*, a human user and an interface agent).
- Effects of automation levels varying from fully autonomous to just suggestive on the collaborative performance with a human user having different skill-levels.
- Effects of parameters of collaborative task environment on the performance; what kinds of cues should be displayed and/or hidden in the interface for the collaboration.
- Characteristics of collaboration life cycle on how the social relationships between a human user and an user-adaptive interface agent do vary as a human improves his/her control skills.

We believe that the key for the success of socially-centered automation is to develop an inferential component of the agent on evaluating the quality of interactions made between the human and the agent quantitatively. The above Lens Model Analysis is one of the promising frameworks for this purpose. We are also analyzing the actual data of human-agent interactions and we have shown how the relationships between the human and the agent does evolve with respect to the indices as their collaboration and concept-sharing proceed along the time horizon of interactions (Sawaragi *et al.*, 2000; Sawaragi 2001; Sawaragi *et al.*, 2001a).

We apply the above Lens Model analysis to evaluate the human-agent interactions. For this purpose, we expand the original Lens Model relating between the human and the environment (*i.e.*, criteria). This is explicating a fact that there are two *independent* judges made by a human user and by an artifact agent. This newly derived Lens Model seems to be similar to the original Lens Model, but instead of evaluating the degrees of coupling strength between the human and the task environment, this new model can evaluate the degrees of coupling strength between the human and the agent from the accumulation of the ongoing behaviors by the two actors (*i.e.*, a human and an agent).

We presents a number of indices derived from the analysis of interactions using this modified Lens Model and proposing a method for quantifying the degrees of sociality realized between human users and agents.

- *agreement r_a*: an index of the matching between the judgments made by the human and the agent
- *explicit knowledge shared G*: an index of the degrees of concept sharing with respect to their policies (*i.e.*, degrees of sharing explicit knowledge)
- *cognitive control R_h and R_a*: indices for measuring how the human and the agent are making their respective judgments according to their respective policies
- *tacit knowledge shared C*: an index of the degrees of concept sharing with respect to what are not captured by the linear policies (*i.e.*, degrees of sharing implicit knowledge)

Currently, we are analyzing the dynamics of sociality and investigating into how the interactions change when a user's proficient skill levels as well as an agent's automation levels are varied.

5. CONCLUSIONS

In this paper, after surveying the problems incurred by the conventional technology-centered automation in a variety of fields, we put an emphasis on the fact that a concept of sociality is really needed to form the ideal relations of human-automation and to let them emerge out of intimate interactions. Then, we presented an *interface agent* as a semi-automated artifact, and discuss about its social relationships with a human user. Although the details of our currently ongoing works were not shown due to the lack of space, they will be presented in the oral presentation opportunity.

REFERENCES

Billings, C.E. (1997). *Aviation Automation: The Search for a Human Centered Approach*, Lawrence Erlbaum Ass.

Cooksey, R.W. (1996). *Judgment Analysis: Theory Methods and Applications*, Academic Press.

Sheridan, T.B. (1997). Human-Centered Automation: Oxymoron or Common Sense?, *Proc, of IEEE Int. Conf. on System, Man and Cybernetics*, Vancouver, Canada.

Maes, P. (1994). Agents that Reduce Work and Information Overload, *Communications of the ACM*, **37**-7, pp.30-40.

Sawaragi, T., Ishibashi, A., Nakatani, Y., Kawai, K. and Katai, O. (1999). Coordination Quality in Automation, *Post Conference Proceedings of the 7th IFAC/IFIP /IFORS/IEA Symposium on Analysis, Design and Evaluation of Man-Machine Systems*, Pergamon Press.

Sawaragi, T. (2001). Lens Model Analysis of Interactions between Human and Interface Agent for Realizing Open, Social and Usable Automation Systems, *Proc. of the 8th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems,* Kassel, Germany, September 18-20.

Sawaragi, T., Horiguchi, Y. and Suzui, Y. (2001a) Co-Adaptation of Human and Machine Autonomies Mediated by Ecological Environment, *Proc. of Human Computer Interaction International,* New Orleans.

Sawaragi, T., and Shimokawa, T. (2001b) Semiotic Learning for Social Robot Using Evolutionary Computing, in Baba, N., Jain, L.C. and Howlett, R.J. (Eds.), *Knowledge-Based Intelligent Information Engineering Systems & Allied*), pp.530-534.