《Research Paper》

Journal of the Human Engineering Society of Korea
Vol. 6 No. 1, April. 1987

AIDING THE OPERATOR DURING NOVEL FAULT DIAGNOSIS

Wan C. Yoon and John M. Hammer

Center for Man-Machine Systems Research
Georgia Institute of Technology
Atlanta, Georgia 30332

ABSTRACT

The design and philosophy are presented for an intelligent aid for a human operator who must diagnose a novel fault in a physical system. A novel fault is defined as one that the operator has not experienced in either real system operation or training. When the operator must diagnose a novel fault, deep reasoning about the behavior of the system components is required. To aid the human operator in this situation, four aiding approaches which provide useful information are proposed. The aiding information is generated by a qualitative, component-level model of the physical system. Both the aid and the human are able to reason causally about the system in a cooperative search for a diagnosis. The aiding features were designed to help the human's use of his/her mental model in predicting the normal system behavior, integrating the observations into the actual system behavior, or finding discrepancies between the two. The aid can also have direct access to the operator's hypotheses and run a hypothetical system model. The different aiding approaches will be evaluated by a series of experiments.

INTRODUCTION

In highly automated systems, the human operator is primarily a monitor and supervisor [Rasmussen 1983, 1984]. An important monitoring function is diagnosing equipment faults, a difficult task in automated systems. The current approach to fault diagnosis is to train the operator to deal with relatively common faults. The training might teach the operator to use symptoms to distinguish faults and to follow procedures to correct them. While this approach should be successful with common faults, it does not support diagnosis of novel faults.

Recently, there has been much interest in supporting the human operator via expert systems for diagnosis. To be sure, this approach will improve the system performance on relatively common failures. As for novel failures, many expert systems for diagnosis [Shortliffe 1976, Miller, Pople, and Myers 1984] are based on shallow reasoning: a set of symptoms suggests a diagnosis. This mapping is not explicitly based on a system model. Consequently, such systems are subject to the same limitations as training and procedures. The designer may have to anticipate the failure for the expert system to solve it correctly.

Aiding from Deep Reasoning

In contrast to the above, our aid is based on deep, causal reasoning about the system. There are several advantages to this approach. First, novel fault diagnosis is normally considered to be knowledge-based reasoning [Rasmussen 1983]. Hence, it seems appropriate for an intelligent aid to reason causally. Second, this approach should be more reliable and robust. The system knowledge is represented at the component level. Because components are small and comprehendable, it should be possible to create representations that are correct, perhaps even provably so. These points support the belief that causal reasoning can cover a wider range of faults [Davis 1984].

In spite of the power of the intelligent aid, we believe there are several reasons to keep the human in command of the problem solving. First, diagnosing a novel failure may require the human to extend the aid's model. Second, when diagnosis involves operating the system (e.g., opening valves, starting motors), it would be better to leave these operations to the human. Third, it may be that the human and the aid may be better able to find a solution cooperatively than either can alone. This is possible, even necessary, for two reasons. The human has better pattern recognition capabilities and can make inductive leaps. Second, the human may need to resolve ambiguities inherent in the aid's model.

Suboptimalities in Human Problem Solving

The aid is designed to mitigate human suboptimalities that occur during decision-making and troubleshooting [Wickens 1984]. Two categories of suboptimalities used here are knowledge-limited and cognition-limited. The knowledge-limited suboptimality is simply that the operator does not fully understand the system. Obviously, the aid's model is a basis for compensating for this problem. There are many cognition-limited suboptimalities. The required information processing for a deep-reasoning diagnosis of a complex system can overload the operator's limited mental resources (i.e., attention and working memory). To help, the computer aid can process and display useful information for the operator. We expect that this may mitigate the human biases in two ways. First, when the human relies on the aid for some stages of his problem solving, the information received from the computer is not biased. For example, if the human uses the aid to test a hypothesis, the confirmation bias (i.e., the tendency to seek only confirming evidences) will be prevented since the computer is not susceptible to this bias. Even if the human does not rely on the aid, it may be able to display information that is compatible with the human's processing. Bias may be avoided if the operator compares his results to the aid's. Second, the aid provides some of the needed information processing, the human is to free to concentrate on other areas. For example, if the workload of hypothesis evaluation is lessened by the aid, the human may more freely entertain different hypotheses rather than stick to one hypothesis (i.e., anchoring).

In the subsequent sections of this article, we will review some relevant research on novel fault diagnosis, discuss the context of our experimental task, and discuss the qualitative model in our aid and its expected effects.

REVIEW OF NOVEL FAULT DIAGNOSIS IN COMPLEX SYSTEMS

The literature on novel fault diagnosis in complex systems is limited. The section will have three parts. The first is empirical research on the effects of training on diagnosis. The second is Rasmussen's system engineering approach to the information needs of operators. The third is Wohl's performance model for predicting diagnostic times for novel failures. The last is the human information processing view of problem solving, which is similar in some ways to novel fault diagnosis.

Shepherd et al. [1977] have studied the effects of training on the errors operators committed while diagnosing familiar and unfamiliar failures. There were three kinds of training. The first was "no story," which amounted to a brief introduction to the control panel instruments. The second was "theory," in which the operation and flow of materials was explained. The third was "rules," which included the above theory training plus a set of proceduralized rules for diagnosing failures. After this training was administered, the three groups were tested. All three groups were significantly different, with rules best and the no story group worst on accuracy. The groups were then trained by examples to diagnose faults, and a second test revealed no differences between the groups. Later, all groups were tested again with two sets of faults -- familiar and unfamiliar. Familiar faults were diagnosed equally well by all groups, but unfamiliar faults were diagnosed best by the rules group.

An experiment on the effects of training on operator control of a simulated process control plant has been conducted by Morris and Rouse [1985a]. One situation examined was the diagnosis of novel failures for which some of the subjects had sufficient theoretical training to diagnose the failure.

The system controlled was a network of fluid tanks. Fluid was pumped from these tanks through valves to neighboring tanks. Two novel failures were studied: a tank rupture that caused a loss in fluid, and a safety system failure that caused the system to shut down when it was not in danger. The experimental results did not show any differences due to training. Nearly all subjects were able to diagnose the tank rupture, and only half were able to diagnose the safety system failure.

System Engineering and Complex Diagnosis

Rasmussen [1983] has discussed operator control of complex systems in terms of three levels of information processing: skills, rules, and knowledge. Skill-based performance applies primarily to automatic, sensory-motor tasks that proceed without conscious control. One characteristic of such performance is that it is not decomposable or verbally expressible (for example, one cannot verbalize the skill of riding a bicycle).

The rule-based level is the second level of processing. A rule is a direct mapping from a set of input symptoms to a diagnosis or action. While performing at this level, the operator does not make recourse to causal models. Rule-based reasoning can be verbalized, which distinguishes it from the previous level.

The knowledge-based level is most relevant to the research reported here. Knowledge-based reasoning must be applied when novel failures occur.

Neither skill-based or rule-based behavior should be used, and hopefully, the operator realizes this (but there is no guarantee). The operator's control occurs by first forming a goal and then a plan consisting of actions that lead to the goal. The plan is evaluated and perhaps modified by a combination of mental simulation or actual actions taken on the machine. Mental simulation relies, among other things, on the operator's mental model of the system.

Rasmussen [1985] has discussed functional and causal reasoning in diagnosis and control of complex plants during novel failures. Physical systems may be represented along a hierarchical, causal-functional continuum. The causal end of this dimension is a description of components according to their local behavior and their physical and structural location (much like a qualitative model). The functional end of the dimension is a description of aggregates according to their function or purpose. In highly automated systems, the operator also needs to know the intent of the automation, since it can change both the function and structure by its own action. The implications for novel fault diagnosis are the claims that an operator needs a multilevel display for intention, function, and causation. The motivation for this is that diagnosis begins at a functional level and moves toward a causal level as the diagnosis becomes more precise.

Maintenance Complexity

Wohl [1982] has observed that electronic troubleshooting in complex equipment operates in two modes. This first mode is for routine failures, which account for 65-80% of all failures. These are repaired relatively quickly. The second mode is for novel failures, which require substantially more time to diagnose and lengthen substantially the mean time to repair. A model for predicting the frequency distribution of novel malfunction repairs has been developed and tested. The model has three parameters: an equipment complexity index, which is the average connectivity of a component; second, an average time to test a component; and third, a parameter that describes how diagnostic interpretation becomes geometrically more complex with each diagnostic test. The test of the model showed a correlation of r=.98 between measured and predicted mean time to repair for fourteen different electronic systems. In a related article, Wohl [1983] observed that the model predicted an infinite mean time to repair when the equipment complexity index exceeded 7.5. An infinite mean time to repair simply means that some malfunctions are never diagnosed. An equipment complexity index of 7.5 means that the average component is connected to 7.5 other components. This limiting value is close to the chunk capacity of human working memory. This result is consistent with the often observed relationship between connectivity and diagnosis complexity.

Complex Diagnosis and Human Problem Solving

Much of the research on problem solving would appear to be relevant to novel fault diagnosis [Newell and Simon 1972]. We briefly review here the human information processing approach to modeling of problem solving and then discuss how novel fault diagnosis differs from it. The information processing approach is centered around the idea of a problem space, which is the human's representation of the key characteristics of a problem. The subject is given an initial and goal state in the problem space and a set of operators that transform the problem from one state to another in the problem space. Usually, the states and operators are crisply defined. Often, there is a metric for the difference between a given state and the goal state. This metric can be used as a heuristic for selecting the operator that moves the greatest distance toward the goal.

The behavior of a human is modeled by a production rule system. Each production rule contains a condition and an action. The condition is a boolean expression on the features of the problem space, some of which are in the human's working memory and some of which are externally perceivable. The potential actions are working memory changes or operators as described above.

Clearly, novel fault diagnosis is a special case of problem solving. The specializations are as follows. First, the human operator must realize the presence of a novel rather than routine failure. Ideally, the displays that result from a novel fault would be sufficiently different from the displays of routine faults. If the novel fault had a display different from routine faults, detection of a novel fault would seem to be assured. Unfortunately, no existing system has been designed from this perspective.

Another specialization is that novel fault diagnosis will occur when the operator has a problem space designed for routine operations and routine failures. It is not known if an existing problem space representation will interfere with novel fault diagnosis. It would seem difficult to believe that some interference does not occur.

A final distinction between novel fault diagnosis and most problem solving research has been how clearly the human can observe the system and the consequences of changes to it. For example, in cryptarithmetic, the human has complete information about the system, the legal operations, and their immediate consequences. Typically, when an operator controls a complex system, the system state is less clearly perceived, the available operations are larger in number, and their effects less clearly perceivable. The consequences of this imprecision are not well understood.

THE SYSTEM AND THE TASK

The Orbital Refueling System (ORS), a NASA-designed payload on the Space Shuttle, was selected for study [NASA 1985]. The function of the ORS is to refuel orbiting satellites with hydrazine, with the objective of extending their useful service life. As shown in Figure 1, the ORS fluid system contains a variety of components such as tanks, valves, pipes, etc. The operator controls the simulated ORS by opening and closing valves. Transferring fuel from propellant tank 1 to propellant tank 2 might proceed as follows. First, tank 2 pressure is reduced by momentarily opening valves 10, 11, 13, and 17. Second, tank 1 is pressurized by opening valves 1, 3,

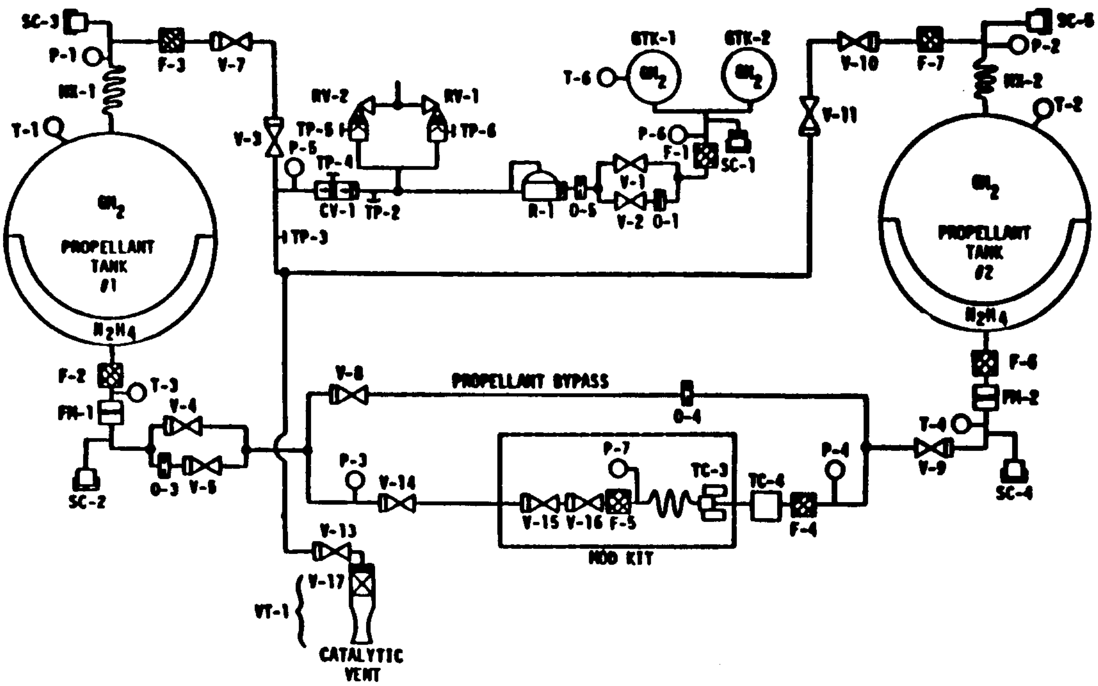


Figure 1. The Orbital Refueling System.

and 7. Gaseous nitrogen will flow out of the two small supply tanks, be pressure regulated, and fill tank 1 on one side of the bladder. To transfer fuel to tank 2, valves 5, 14, 15, 16, and 9 would be opened. Because this version of the ORS was for demonstration purposes, all transfers take place between the two large tanks rather than to a satellite fuel tank. There are several assemblies whose purpose was not explained in the above example. The relief valves RV1 and RV2 serve as a safety pressure relief. Check valve CV1 prevents backflow into the gas system. The bladders in tank 1 and 2 serve to isolate the fuel from the propellant and also to contain the fuel in the weightlessness of space. Some components (e.g., valves 10 and 11) may seem redundant; they are so by design for two failure tolerance.

The Diagnosis Task

The operator's task is to diagnose the failure in the system. This requires the operator to manipulate and observe the system, because a diagnosis cannot be determined uniquely from an observation of a state vector at a single point in time. A solution is an assignment of states to components such that the assignment's behavior is always identical to system behavior. For a single valve failure, the solution would be a normal state for all components save the failed valve, which might be jammed shut. The diagnosis problem can be viewed as a combinatorial search for a state assignment. The search is constrained by the laws of component physics. That is, a state assignment to a component imposes constraints on its neighboring components. For example, if a valve is opened and permits a flow down a pipe, the component receiving the flow must be in a state to accept the flow.

QUALITATIVE MODELS OF CONTINUOUS PHYSICAL PROCESSES

This section describes qualitative models: representations, the computational problems solved, and the specific needs of our aid of the qualitative model.

A qualitative model is a symbolic representation of a system. Its most basic description is of a component. A component is described in terms of its connections to other components and its behavior. Behavior is described in terms of the physical variables which are present at its connections. The differentiation between the structural description (connections) and the behavioral description is particularly important for insuring the robustness of a qualitative model. The isolation of each component in the behavioral description has usually been emphasized by other qualitative modeling [De Kleer and Brown 1983]. Contrarily, our qualitative model represents the system at both the component level and at an aggregated level as paths. The motivation for this is the belief that a multi-level description is closer to the operator's internal model of the process. In fact, more effective communication between our model and the human operator was enabled by the use of the higher level description.

From a given state, the behavior of a component is described in terms of the physical variables present at its ports. A physical variable (and its time derivative) may take several values. The time derivative usually has only one of three possible values: negative, zero, or positive. The variable itself may take either nominal or ordinal values. The nominal values usually correspond to points at which behavior (component or material) changes. For example, water temperature would have nominal values at freezing and boiling. Variables may also take on ordinal values (or relationships). For example, water temperature could be taken to be greater than freezing and less than boiling.

The nominal and ordinal values taken by physical variables are said to occur in a <u>quantity space</u> [Forbus 1984, Kuipers 1984]. The quantity space is a partial ordering on the physical variable values it contains. The partial ordering occurs because not all comparisons are relevant to understanding the physical system qualitatively. For example, consider a valve between two tanks, A and B. When the valve is opened, the resulting behavior is determined by the pressures in two tanks. The pressure at other unconnected points in the system is unrelated to the above behavior.

AN EXPLORATORY EXPERIMENT

An exploratory experiment was conducted to observe the strategies subjects used to diagnose the ORS. Three Georgia Tech students were used as subjects. The use of college students is usually considered a compromise in experimental research. Since some space shuttle astronauts have been engineers, this compromise is reasonable in this situation.

The training contained both theoretical and practical elements. First, the basics of gas and fluid transfer were reviewed. Second, there was an explanation of the normal and malfunction behavior of each component. Third, subjects were told how to test for a failed component and how to operate the system.

The subjects then solved five single failure malfunctions. The failures included two leaking valves, a ruptured pipe, a pressure transducer that always read high, and a relief valve that opened inappropriately. The data collected included a time-stamped record of the ORS commands issued and a tape recording of the subject's verbal protocols [Ericsson and Simon 1984]. The time to solution ranged from 3.6 to 31.1 minutes showing an average of 14.2 with a standard deviation of 8.2 minutes.

A Post-hoc Analysis of Performance Data

The data from our preliminary experiment suggest several interesting characteristics of human diagnosis behavior, and which in turn suggested some directions for computer aiding. First, the time spent for a successful diagnosis is strongly related with the number of information gathering actions (IGA) ($r = 0.79$) and the average time between actions ($r = 0.77$). The latter two variables were not strongly correlated ($r = 0.21$). The implication of this is reducing the number of information gathering actions (IGA) is an important goal for improving diagnostic performance.

Second, we classified IGA's into effective ones (EIGA), which reduced the size of feasible hypothesis set, and ineffective ones (IIGA), which did not. We found that the number of EIGA is invariant among subjects and is also not significantly correlated with the total number of IGA. The total number of IGA is correlated with IIGA (corr.= 0.98), which outnumbered EIGA by 2.5 : 1. This suggests that a problem is solved by collecting the right number of EIGA (largely determined by the complexity of the problem). A better performance is possible when the effective actions are executed earlier in the diagnosis.

Third, we investigated how well the subjects detect the abnormal behavior of the system. We assessed the delay in diagnosis due to failures to collect information that would have revealed the abnormal system behavior. The delay showed high correlation ($r = .79$) with the number of ineffective actions. Also, 75% of effective actions were of abnormal behavior, and the remaining 25% were of normal behavior (negative evidence). Observations on abnormal behavior, if they are correctly interpreted, became effective actions in almost all cases. Thus, abnormal behavior of the system is probably the most important source of effective information.

The conclusion is that, to help the diagnosis, the cues for effective actions need to be given. Abnormal system behavior is worth watching for this purpose. When designing an aid, a major advantage of using abnormal behavior is that inferring or requesting the human's current hypothesis is not necessary.

Observation of Strategies

There appeared to be three strategies that subjects used: hypothesis-driven evaluation, data-driven evaluation, and topographic search. Hypothesis-driven evaluation starts with the planning of a test procedure for a given hypothesis. The hypothesis needs to be explicit enough to enable the prediction of its resulting system behavior. A test plan would be diagnostic if, given that the hypothesis is true, the response of the system to the test is unique to the hypothesis. When a sufficiently diagnostic test has been planned, the test is executed and its result evaluated.

This evaluation tends to be short because it has already been determined what the results might be.

With data-driven evaluation, the subject first examines a piece of data to determine if it is worth closer attention. This examination is done by comparing the data to expected system behavior. If the data turns out to be unexpected (i.e., not explained in terms of previously observed symptoms or normal behavior), then hypotheses are formulated to explain the data. Whether the formulation is successful or not, this piece of data is remembered by the diagnoses as another symptom to be used later during diagnosis.

Topographic search seems to help reduce the mental workload in diagnosis. Both above evaluation strategies involve deep reasoning with functional causalities. With deep reasoning, the former deduces necessary data from a given hypothesis while the latter formulate and evaluate hypotheses from the given data. Topographic search [Rasmussen 1984], without such a deeply based hypothesis, is used to find data. For instance, the sensor near the suspected component are read in hope that the reading may give some diagnostic information. An example of topographic search of hypotheses is suspecting nearby components when a sensor reading is out of the normal range. The differentiation of a single general hypothesis to several more specific hypotheses can be considered as topographic search.

AIDING WITH A QUALITATIVE MODEL

This section describes how the qualitative model is used as a foundation for aiding. First, each window of the interface will be described. Four different aiding strategies and the motivation for each of them will then be presented. Each strategy emphasizes different type of aiding information.

ORS Interface

The interface has four windows: schematic, interaction, sensor display, and hypotheses (Figure 2). The schematic window displays a schematic diagram of the ORS. The schematic always shows the commanded state of the valves. First, tank 2 pressure is reduced by momentarily opening valves 10, 11, 13, and The interaction window is where the operator's commands are echoed by the interface. The commands available to the operator include the following:

(1) Opening and closing valves.

(2) Comparing two pressures. On a real physical system, the numerical pressure could be displayed on the schematic. When a qualitative model is used to simulate the physical system, there is no absolute scale in general to which a pressure can be referred. Instead, a pressure can be compared to other pressures in the system by the relations less-than, equal-to, or greater-than.

(3) Display of the first derivative of a pressure (positive, zero, or negative).

And, when the corresponding aiding feature (it is described more fully in a later section) is available,

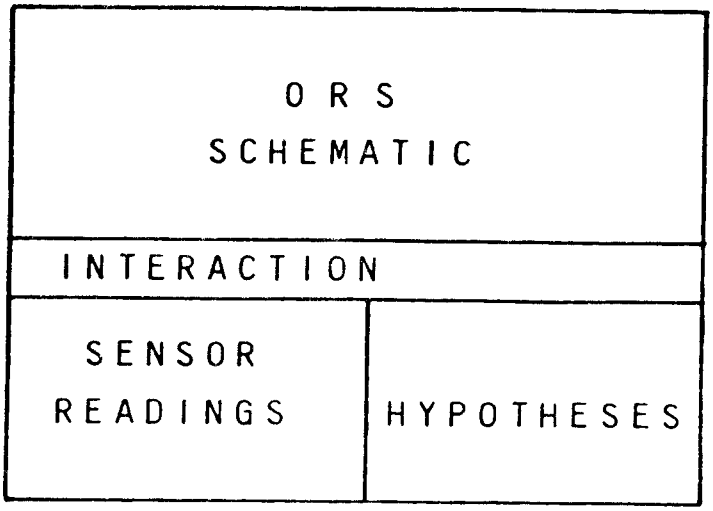


Figure 2. The operator's display.

(4) Turning the what-if model on and off.

(5) Making state assumptions in the what-if model.

The sensor display contains the output from the sensor display commands: the relationship between two pressures or the first derivative of a pressure. When appropriate aiding features are activated, suggested sensor readings will also be displayed in this window.

The hypotheses window displays a set of hypotheses that are set by the operator. These hypotheses are simply state assignments to components (e.g., valve 13: leaking). Pipes, which do not have names displayed in the schematic, are designated as left or right to named components such as valves and orifices. For example, the pipe between valve 8 and orifice 4 is designated either (r v8) or (l o4). The operator can freely add or delete the hypotheses.

Aiding Approaches

Based on observed human strategies of diagnosis, four aiding approaches seem to deserve evaluation. Each approach emphasizes different information and uses an appropriate communication mode for the kind of information.

Topographic Aiding. The first and second aiding approaches are based on two presumed forms of operator cognitive processing. First, the operator must observe and infer what the system is actually doing. This processing is termed O (Observed) and is concerned with flows, leaks through valves, leaks out of pipes, and the general vicinity of the fault. Second, the operator needs to generate normal system behavior to compare with observed behavior. This processing is termed N (Normal). Two obvious forms of aiding are to generate O and N so that the operator does not have to devote cognitive processing to generating them. To produce O, the aid integrates the information from the pressure sensors to which it has continuous access.

Like a human operator, the aid has to guess the actual behavior from the sensor information since it cannot know the real system state. In contrast, N is generated by the qualitative model under the assumption that every component obeys the command.

O and N are displayed topographically. For both O and N, the aid displays two forms of system behavior: equal pressure paths and mass flow paths. The former is the set of components that should be at equal pressure given the commanded valve positions. Whenever the operator creates an equal pressure path by opening a valve, the path is highlighted. Similarly, a mass flow path created by an operation is highlighted as long as it exists.

Figure 3 is an example of N display. Opening valve 9 was the latest change. This would make, if the system were fault-free, the pressure equal through the highlighted path. First, tank 2 pressure is reduced by momentarily opening valves 10, 11, 13, and First, tank 2 pressure is reduced by momentarily opening valves 10, 11, 13, and

Figure 4 shows the same configuration as Figure 3, except that the O display (rather than N) is activated. When valve 9 was opened, the pressure p2 began to decrease and p1 increase. This leads the aid to believe there is a mass flow from tank2 to tank1 (the path is highlighted) in spite of the closed positions of valve 8 and valve 15. However, since the aid cannot be certain which valve is leaking, it highlights both paths. When a precise conjecture is not possible, the aid will take a conservative position as in this example. Quite naturally, O and N aiding cannot be used simultaneously.

<u>Differencing Observed and Normal Behavior</u>. The third aiding approach is to suggest observations that reveal the differences between the observed system behavior and the normal system behavior. This difference will be

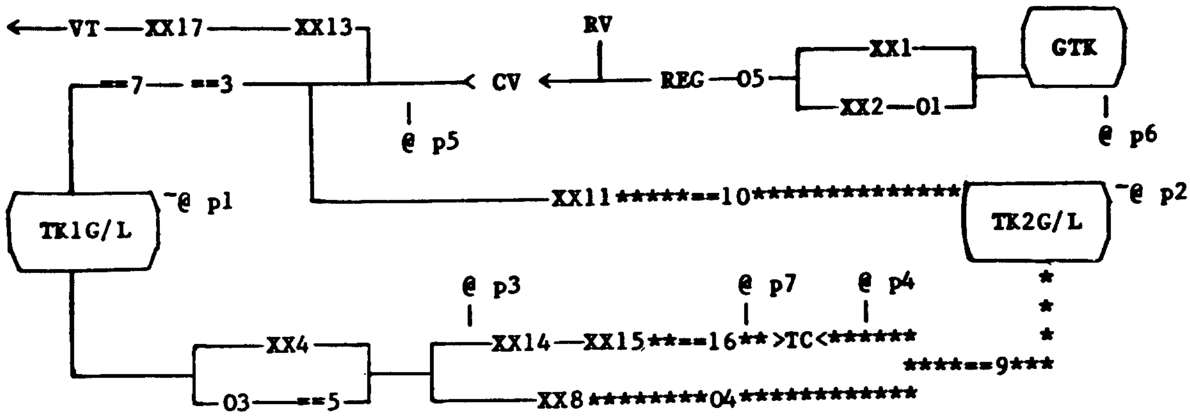


Figure 3. The normal response (N).

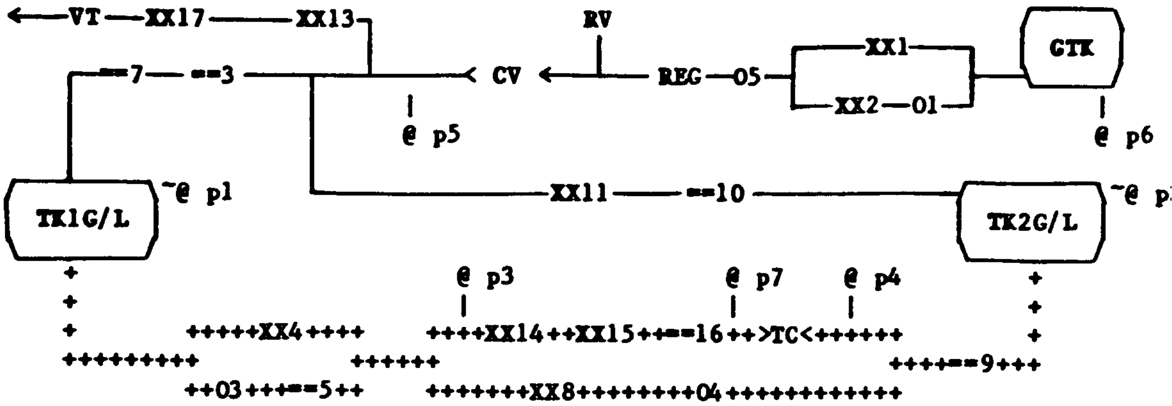
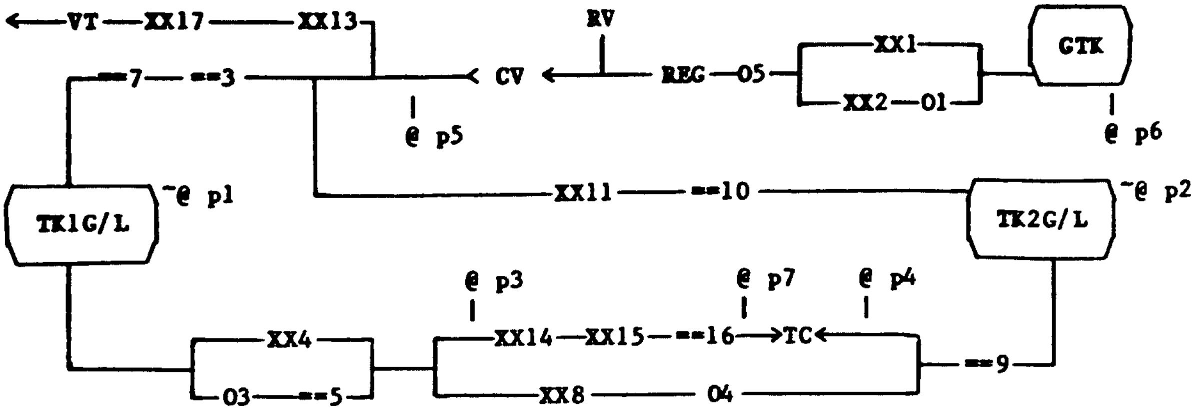


Figure 4. The observed response (O).

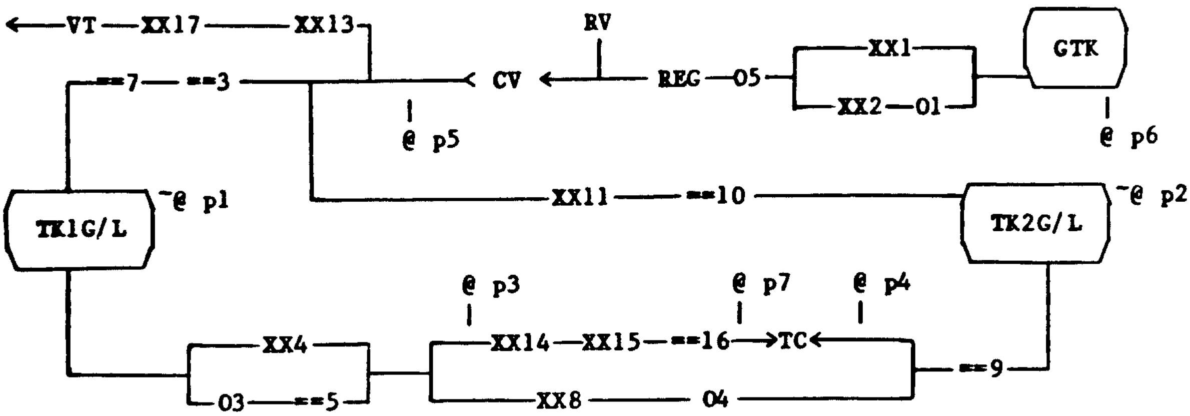
connection with the results of our preliminary experiment. Such a deviation from normal behavior, when observed and correctly interpreted, helped effectively reduce the size of the feasible hypothesis set. Figure 5 shows an example of this feature's display in the same situation as of Figure 3 and 4. The aid suggests, for example, to issue a command (i pl), which is to inquire the first derivative of pl. When the operator follows this, he will find pl is increasing, which is opposite to the commanded situation (no flow should be possible from either GTK or TK2G/L). First, tank 2 pressure is reduced by momentarily opening valves 10, 11, 13, and

<u>The What-if Model</u>. The fourth, and the last, aiding feature is closely related to the above. This feature can use any hypothetical behavior (denoted by H), instead of the normal behavior, with which to contrast the observation. The operator can freely set or remove hypotheses. Then, the aid will run a what-if model based on the hypotheses in place of the normal model. Any discrepancies (denoted by O-H) will be reported in the same way. If the operator's hypothesis is a leak in v10, the feature would present a display shown in Figure 6. First, tank 2 pressure is reduced by momentarily opening valves 10, 11, 13, and Note that if no hypothesis is stated, the recommendations would be the same as the previous example (i.e., O-H = O-N if H = N). Further, if the hypothesis is incorrect, the aid will recommend readings. If the hypothesis is correct, the aid will be silent.



See
* i p1
* i p2
* c p2 p4

Figure 5. Deviation from normal behavior (O-N).



(1 v10) : 1k

See
* i p1
* c p2 p4

Figure 6. Deviation from hypothetical behavior (O-H).

Comments

The first question is which of O and N aiding (they are mutually exclusive) will work better. The answer should be related to the human strategies and other characteristics of the human's information processing.

Second, is O-N, which is more specialized and explicit, better than O or N, which are more general? The advantages of the latter are their generality, topographic presentation, and hence, easy communication. Their disadvantage is that they require interpretation and do not direct the operator to take a specific action.

Third, effectiveness of O-H depends on system complexity. With high complexity, the disadvantage of explicit hypothesis communication may be offset by the complex calculations the aid can do for the operator. This feature will be evaluated only on one system, so complexity effects cannot be measured. Therefore, the evaluation of this feature will depend on its correct use rather than a performance improvement.

The common motivation for these aiding approaches is to perform computations that the operator is believed to make when diagnosing the system. As much as these computations are related to the human's mental model, the qualitative model in the aid may be an appropriate vehicle to help or replace the computations. There are two ways this approach might help. First, the operator may have an incorrect or incomplete mental model. Second, the operator may have difficulty integrating correct component behavior into correct system behavior because of cognitive limitations. The aiding approaches support different uses of the mental model: to envision the normal or hypothetical behavior, to conjecture the actual behavior, and to describe the difference between behaviors of two (e.g., O and H) models. This does not mean the operator need not understand the system at all; he or she still needs to understand the meaning of aid's information and select the hypotheses.

CONCLUSION

An aid has been described for novel fault diagnosis in complex systems. To the best of our knowledge, this aid is unique in the following ways. First, the emphasis is on novel rather than routine faults. Second, it contains a qualitative model that may correspond to the human's internal model of the system. This model represents knowledge only of how the system works. Many of the proposed aiding schemes are proceduralized fault finders: they tell the operator what action to take. Third, the qualitative model is the basis for much of the aiding that takes place. Fourth, the interface specifically attempts to mitigate some human decision-making suboptimalities during fault diagnosis.

ACKNOWLEDGMENT

This research was supported by NASA-Ames Research Grant NAG 2-123. Dr. Everett Palmer is the technical monitor.

REFERENCES

Davis, R., "Diagnostic reasoning based on structure and behavior," Artificial Intelligence, Vol. 24, pp. 347-410, 1984.

De Kleer, J. and Brown, J.S., "Assumptions and ambiguities in mechanistic mental models," in D. Gentner and A.L. Stevens (Eds.), Mental Models, Hillsdale, NJ: Lawrence Erlbaum Assoc., 1983.

Forbus, K., "Qualitative process theory," Artificial Intelligence, Vol. 24, pp. 85-168, 1984.

Gentner, D. and Stevens, A.L. (Eds.), Mental Models, Hillsdale, NJ: Lawrence Erlbaum Assoc., 1983.

Kuipers, B., "Commonsense reasoning about causality: Deriving behavior from structure," Artificial Intelligence, Vol. 24, pp. 169-203, 1984.

Mehle, T., "Hypothesis generation in an automotive malfunction inference task," Acta Psychologica, Vol. 52, pp. 87-106, 1982.

Miller, R.A., Pople, H.E. Jr., and Myers, J.D., "INTERNIST-1: An experimental computer-based diagnostic consultant for general internal medicine," in Readings in Medical Artificial Intelligence, Clancey, W.J. and Shortliffe, E.H. (eds.), Reading, MA: Addison-Wesley, 1984.

Morris, N.M. and Rouse, W.B., "The effects of type of knowledge upon human problem solving in a process control task," IEEE Trans. Systems, Man, and Cybernetics, Vol. SMC-15, No. 6, 1985.

Morris, N.M. and Rouse, W.B., "Review and evaluation of empirical research in troubleshooting," Human Factors, Vol. 27, No. 5, October 1985.

NASA Johnson Space Center, "Orbital refueling demonstration system description," Program Development Office, October 21, 1985.

Newell, A. and Simon, H.A., Human Problem Solving, Englewood Cliffs, NJ: Prentice-Hall, 1972.

Rasmussen, J., "Skills, rules, and knowledge: signals, signs, and symbols, and other distinctions in human performance models," IEEE Trans. Systems, Man, and Cybernetics, Vol. SMC-13, No. 3, pp. 257-266, 1983.

Rasmussen, J., "Strategies of state identification and diagnosis in supervisory control tasks, and design of computer based support systems," in W. Rouse (Ed.), Advances in man-machine systems research, Vol. 1, JAI Press, 1984.

Rasmussen, J., "The role of hierarchical knowledge representation in decision making and system management," IEEE Trans. Systems, Man, and Cybernetics, Vol. SMC-15, No. 2, pp. 234-243, 1985.

Rouse, W.B. and Morris, N.M., "On looking into the black box: Prospects and limits in the search for mental models," to appear in Psychological Bulletin, 1986.

Shepherd, A., Marshall, E.C., Turner, A., and Duncan, K.D., "Diagnosis of plant failure from a control panel: A comparison of three training methods," Ergonomics, Vol. 20, No. 4, pp. 347-361, 1977.

Shortliffe, E.H., Computer-Based Medical Consultation: MYCIN, New York: American Elsevier, 1976.

Wickens, C.D., Engineering Psychology and Human Performance, Columbus, OH: Charles E. Merrill, 1984.

Wohl, J.G. "Maintainability prediction revisited: Diagnostic behavior, system complexity, and repair time," IEEE Trans. Systems, Man, and Cybernetics, Vol. SMC-12, No. 3, 1982.

Wohl, J.G. "Cognitive capability versus system complexity in electronic maintenance," IEEE Trans. Systems, Man, and Cybernetics, Vol. SMC-13, No. 4, 1983.