

## AN EXPERT SYSTEM FOR FAULT MANAGEMENT ASSISTANCE ON A SPACE SLEEP EXPERIMENT

A. ATAMER, M. DELANEY AND L.R. YOUNG \*

*Man-Vehicle Laboratory, Massachusetts Institute of Technology, 77 Massachusetts Ave., Bldg. 37-219,  
Cambridge, MA, 02139, USA*

### INTRODUCTION

At the time of the seminal 1969 Brainerd conference it was evident that engineering analysis tools, as well as engineering hardware, were enabling a revolution in systems physiology. The *hot topic* at that time, reflecting the Messianic faith of Carlo Terzuolo in the omnipotence of engineers, was linear control theory. And indeed, in the subsequent two decades the language and the tools of control theory became part of the lexicon of physiology. Frequency response, feedback stability, non-linear identification, and computer simulation became so entrenched in the systems physiology field that their origins were no longer worth mentioning. The new *hot topic* for physiology became Artificial Intelligence, including learning systems, neural networks, and expert systems.

Our own interests, in MIT's Man-Vehicle Laboratory, evolved from modeling of vestibular and oculomotor function, which was described in Brainerd, to the use of the weightless conditions of space flight for investigation of human spatial orientation. In order to help in the astronaut performance of challenging experiments in space we began to use expert systems for the in-flight guidance of these willing but insufficiently trained surrogate investigators. Principal Investigators can rarely accompany their experiments into space, and ground-to-air contact with the astronauts performing the experiment is not always possible. Consequently, in a typical space flight, astronauts are faced with learning a large number of scientific experiments, in addition to the flight operational procedures. One way to approach both these problems is by provision of a computer decision aid, such as an expert system, to help guide the operator in real time. This paper presents our recent research in the use of such an expert system in the pursuit of a better way of performing space flight experiments.

#### *Background.*

Previous studies have sought to assess the value of expert systems and other computer aids in fault management situations. An expert system designed as a prosthesis was studied by Roth, Bennett and Woods (8). A more controlled study of a computer decision aid for both forced-pace and subject-paced context-independent fault

---

\* Address for correspondence: Dr. L.R. Young, Massachusetts Institute of Technology 77 Massachusetts Ave., Bldg. 37-219 Cambridge, MA 02139, USA. Tel. (617) 253 7759, Fax (617) 258 8111, E-mail: LRY@mit.edu

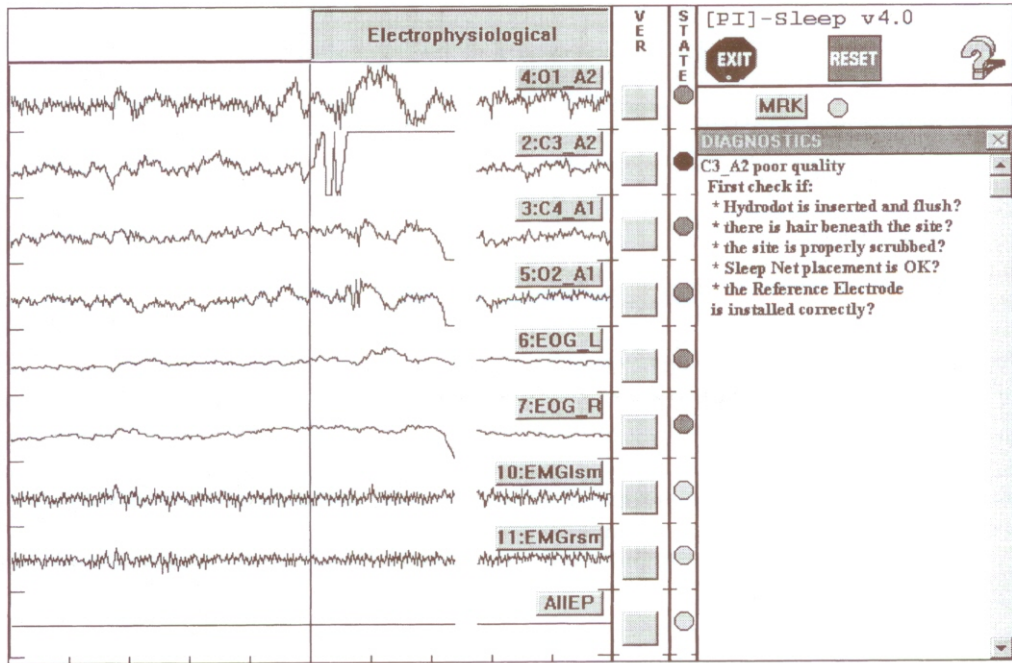


Fig. 1. - *Electrophysiological signals on [PI] interface.*

diagnosis was investigated by Rouse (9). Jones and Mitchell evaluated a software package for helping mission control specialists to perform satellite station-keeping (5). The first experiment in which [PI] was applied was to the Rotating Dome Experiment on STS-58 (4, 12).

A pilot study in 1998 tested subjects' ability to detect poor quality signals (2). Analysis of the pre-sleep data from the Neurolab space flight experiment found that the [PI] indicator lights were correct in 84% of the cases for which it displayed a red signal for non-saturated waveforms. This study also showed that the [PI] indicator lights were found both to reduce the time to detect a problem, and decrease the number of anomalies that went undetected.

The [PI] interface (as shown in Figure 1) consists of a waveform display for sleep electrophysiological signals such as the electroencephalogram (EEG, brain waves), electro-oculogram (EOG, eye movements), and electromyogram (EMG, muscle activity). The interface has two components: an array of indicator lights located beside the waveforms, and a diagnostics window. The indicator lights display [PI]'s assessment of signal quality, and the diagnostics messages show a short version of the NASA malfunction procedures.



## METHODS

The purpose of the ground-based study is to evaluate how much [PI] can improve the astronauts' performance in fault management, such as the time taken to detect a problem ( $T_D$ ), the time taken to troubleshoot the fault ( $T_{TS}$ ), and the percentage of correct detection  $p(T_D)$ .

*Experimental design.*

30 subjects (14 female, 16 male; MIT undergraduate and graduate students) completed all of the testing sessions. The subjects' ability to detect and subsequently troubleshoot faults in the electrophysiological instrumentation was tested both with and without [PI] assistance.

The experiment used a balanced cross-over design, as shown in Table 1. Group 1 had [PI] assistance only on their Day 1 of testing and Group 2 only on their Day 2 of testing. All subjects received the same 3.5 hours of training on the sleep instrumentation and the [PI] interface. Subjects were then tested on two separate days in thirty-minute sessions with the instrumentation.

Table 1. - *Experimental Design.*

	Group 1 N = 14	Group 2 N = 16
Day 1	[PI] ON	[PI] OFF
Day 2	[PI] OFF	[PI] ON

*Apparatus.*

The sleep experiment equipment is comprised of the following:

- 1) Electrode Net (eNet): an elastic web-like cap containing 13 electrode sockets to record the EEG, EMG and EOG signals (Physiometrix, Inc., North Billerica, Massachusetts, U.S.A.);
  - 2) Hydrodots: Ag/AgCl electrodes that fit into the Electrode Net's sockets and contact the skin (Physiometrix, Inc., North Billerica, Massachusetts, U.S.A.);
  - 3) Digital Sleep Recorder (DSR): a device that converts the raw analog signals from the various electrodes and instrumentation to digital signals, which are then recorded onto a PCMCIA FlashRAM card (Copyright 1996 Vitaport EDV System GmbH. Distributed by TEMEC instruments BV, The Netherlands);
  - 4) IBM Thinkpad laptop – Pentium-class processor running Windows 95, with [PI] software installed;
  - 5) Materials for preparing the electrode site (such as swabs, prep solution, adhesive pads, etc).
- [PI]'s knowledge base was developed at NASA Ames using CLIPS, a NASA tool for building expert systems. The [PI]-Sleep software tested here flew aboard the STS-95 mission. Menus for recording the test subject's input were added to [PI] software in the ground study. Also, the color of the indicator light for a questionable signal was changed from amber (as in the flight version) to yellow, because of confusions that arose between the red and amber lights during flight.

*Experimental procedure.*

Each experiment involved the interaction of three people: an MIT research assistant trained by Brigham & Women's Hospital staff, a sleep subject who donned the instrumentation, and the test subject, or astronaut surrogate, who monitored sleep EP signals in real-time. The test subjects were asked to detect a fault, diagnose it, and instruct the assistant how to correct the fault in real time. Interaction was limited: only the research assistants could see or manipulate the sleep instrumentation hardware.

Thirteen trials were carried out on both days. These trials were broken down into three different fault types: null, single-channel and multi-channel faults. Single-channel faults showed only one anomalous channel. Multi-channel faults exhibited more than one anomalous channel. The null fault differed from the other categories because the detection task required subjects to verify that all signals were normal rather than to detect and diagnose a particular fault. The faults appeared in a random order, balanced according to the fault type. The order in which the subjects encountered each trial was reversed on Day 2.

Faults in the system were created, detected, diagnosed, and fixed in real time. A schematic of the experimental setup is shown in Figure 2.

While the assistant created a fault, test subjects were permitted to view the signal display, but wore headphones to prevent audio cues that might hint at the problem they were about to encounter. The research assistant induced a fault in the sleep instrumentation. Once the fault was created, the assistant alerted the test subjects to begin searching for a faulty signal. Test subjects were allotted 180 seconds to detect and diagnose a fault. The subjects removed the headphones and clicked on an event marker, indicating in the data file that they had begun each scenario. They analyzed the signal waveforms and the indicator lights, if available, to determine if a fault existed in the system. If a subject thought a fault existed, he would click on a gray checkbox and select a system state from a list of possible states. If no fault was thought to exist, the subject selected the "System State OK – No Problems" option.

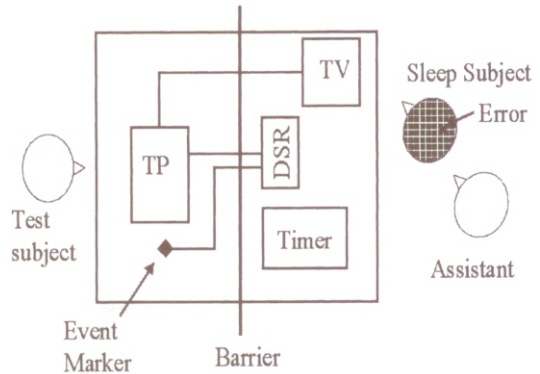


Fig. 2. - Experimental setup.

After assessing the system state, the subject isolated the fault by following the troubleshooting steps outlined, using either [PI]'s diagnostic messages or the actual NASA troubleshooting guidelines given to the astronauts for use in flight. The test subject could question the assistant to gain information about the system, but could not see the instrumentation hardware or interact with it physically. The subject was also trained to ask the sleep subject to make calibration movements, such as "look left, look right," to verify signal presence and quality. Once the fault was diagnosed, the test subject clicked on a gray checkbox to select from a list of possible faults associated with the symptoms recorded earlier. The subject would then ask the assistant to fix the diagnosed fault. After the assistant confirmed that the fault had been removed, the test subject turned back to the first page of the NASA guideline (if used), put on headphones and prepared for the next scenario. Following the experiment, the subjects completed a questionnaire and were debriefed by the assistants.

## RESULTS

The efficacy of [PI] was analyzed in terms of the effect on detection time ( $T_D$ ) and troubleshooting time ( $T_{TS}$ ). A repeated-measures analysis of variance (ANOVA) was performed on  $T_D$ . The Day of the test, a learning effect, was shown to significantly affect  $T_D$  (Table 2). Subjects detected a problem faster on their second day of testing than on their first. Overall, the detection task that took the least time was the multi-

channel fault, then single-channel, then the null faults, as seen in Figure 3. Subjects improved with training from Day 1 to Day 2 for only the null and multi-channel faults. [PI] assistance did not show a significant main effect on  $T_D$ . That is, the status of a signal, provided by [PI]'s indicator lights, proved to have no significant impact on the detection time (3).

Table 2. - Main Effects on  $T_D$ .

Factor	P Value
Day	0.006
[PI]	0.749

A repeated-measures ANOVA was also performed on  $T_{TS}$ . In comparison of group data, the identity of the research assistant who ran the experiment and the group was considered a factor, but did not show a significant main effect. The within-

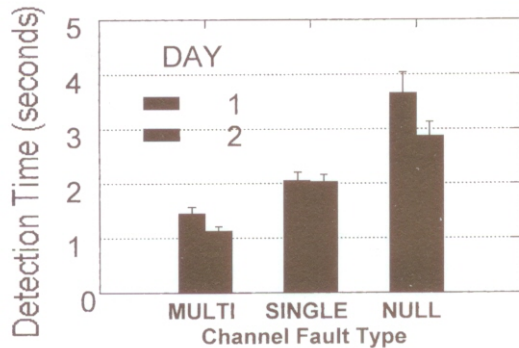


Fig. 3. - Detection time.

subjects analysis shows a significant effect of [PI] on reducing subjects' time to troubleshoot a fault. Training, or Day, had no significant main effect  $T_{TS}$  (see Table 3). These results show that [PI] improved subjects' troubleshooting time, regardless of the day on which they received [PI] assistance. [PI] reduced the troubleshooting time by more than 30% on both days (3).

The effect of [PI] for reducing troubleshooting time can also be seen in Figure 4 (3). The top graph shows Group 1 subjects who received [PI] help on their first day and no help on their second day. All but one of these subjects performed better on their first day with [PI]. The bottom graph shows Group 2 subjects, who received [PI] help on their second day only. All of these subjects performed better on their second day with [PI] help.

The diagnostic reliability of the system is associated with how well people can assess and understand the system state. For this investigation, reliability of fault management was determined via the correctness of fault detection. A detection was considered correct if subjects could find both the correct channel(s) on which the fault was introduced and the correct system state out of all possible states as shown in Table 4 (1).



Table 3. - ANOVA results for Tts.

Analysis of Variance for Tts Between Subjects					
Source	SS	df	MS	F	P
SEX	1643.556	1	1643.556	2.611	0.126
ASST	52.562	1	52.562	0.084	0.776
GROUP	1179.007	1	1179.007	1.873	0.190
Error	10071.413	16	629.463		

Within Subjects					
Source	SS	df	MS	F	P
Day	1041.134	1	1041.134	3.323	0.087
Day*SEX	6.009	1	6.009	0.019	0.892
Day*ASST	8.703	1	8.703	0.028	0.870
Day*GR ([PI])	3872.687	1	3872.687	12.359	0.003
Error	5013.720	16	313.357		

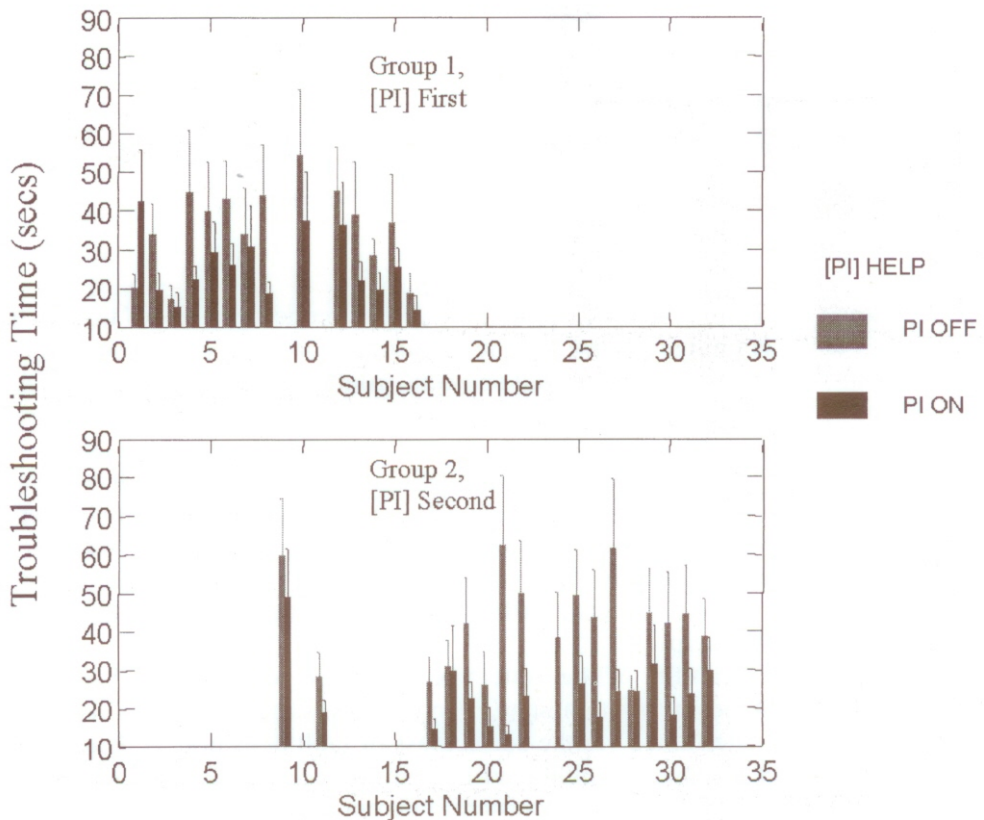


Fig. 4. - Troubleshooting time is significantly improved with [PI] help.

Table 4. - List of System States.

State	Description
1	[PI] signals freeze
2	[PI] displays no signals
3	All EP signals are not present or poor quality
4	EEG signals not present or poor quality
5	EMG signals not present or poor quality
6	EOG signals not present or poor quality
7	System State OK – no problems
8	Other state

percent correct detection rate of [PI] itself was found to be about 60%, as shown in Figure 6 (1).

The data from the pilot study and on-orbit recordings did not assess the subjects' detection correctness as a function of [PI]'s detection correctness. We define the reliability index for the indicator lights ( $t_{fire}$ ) as the time that the indicator remains red for the faulty channel(s) divided by the sum of the time that the indicators are red for any of the channels, whether correct or incorrect.

Subjects were found to be correct on about 80% on their initial detections, as shown in Figure 5 (1). [PI] indicator lights were found to have no significant impact on increasing the percentage of subjects who correctly detected the system state. The reliability of [PI] indicator lights alone was also considered, similar to the observations in Neurolab (2). The

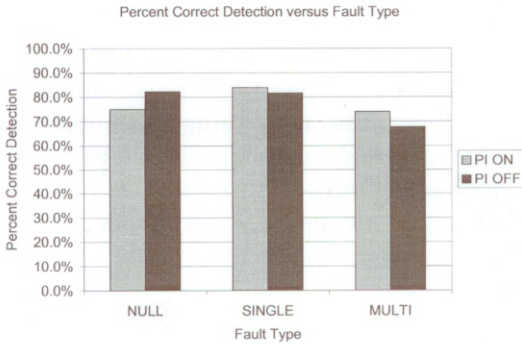


Fig. 5. - Percent correct detection versus fault type.

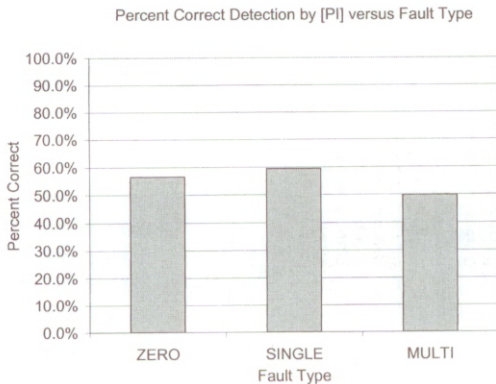


Fig. 6. - Percent correct detection by [PI].

$$\text{Equation: definition of } t_{\text{fire}} \quad t_{\text{fire}} = \frac{t_{\text{ch,correct}}}{t_{\text{ch,correct}} + \sum_i t_{\text{ch,incorrect}}}$$

Only single-channel faults were used in this analysis. Figure 7 is a plot of subjects' percent correct detection rate versus  $t_{\text{fire}}$ , the reliability index of the indicator lights (1). As  $t_{\text{fire}}$  increases, so does the subjects' correct detection rate. The most difficult detection tasks were found to be difficult by both the subjects and [PI]. Subjects' correctness reached a plateau for  $t_{\text{fire}}$  above 0.4. This means that [PI] does not have to be 100% reliable for the astronaut to benefit from the information it provides. On Day 1, subjects with [PI] had a higher chance of correct detection than without it on Day 2, for the most difficult detection tasks. However, subjects with [PI] help on Day 2 (who already had one day's experience with the instrumentation without [PI]) were hindered by the indicator lights; their correct detection rate was lower than that of the subjects troubleshooting without [PI] on Day 2. These strong trends show that subjects in Group 2 had difficulty transferring to an interface *with* an aid from one *without* an aid.

The most significant impact [PI] had on reducing troubleshooting time was in the area of fault identification, which took an unusually long time (i.e. between 120 and 180 seconds). The presence of [PI] reduced the number of undetected anomalies, or timeouts, from 6.34% of trials to 2.27% ( $p = 0.007$ ,  $\chi^2 = 7.37$ ,  $df = 1$ ).

On the questionnaire, 21 out of 30 subjects responded that using both [PI]'s indicator lights and observing waveforms was better than observing the waveforms alone ( $p < 0.0003$ ). Furthermore, 22 people thought that using [PI]'s diagnostic messages was better than using the NASA guideline ( $p < 0.0003$ ). Subjects generally preferred having a checklist rather than consulting the manual. Regarding the reasoning ability of [PI], several subjects indicated that the software should use the system information more effectively by providing an ordered list of the "most probable causes" rather than a generic message list for each problem.

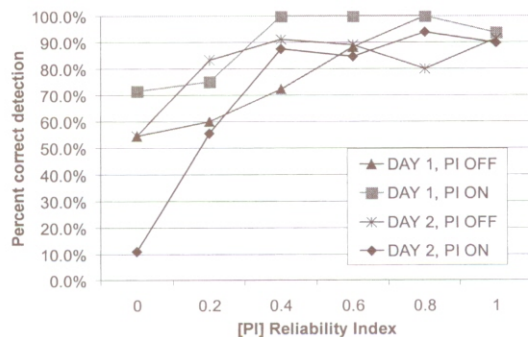


Fig. 7. - Percent correct detection versus [PI] reliability index.



## DISCUSSION

[PI] showed a main effect for reducing troubleshooting time  $T_{TS}$ . One of the driving factors behind this result was that the majority of subjects found the step-by-step troubleshooting help easier to follow on the [PI] software than in the NASA Troubleshooting Guide. One subject commented, "The [PI] diagnostics were much easier to use than the NASA troubleshooting guide, probably because the computer was already in front of me." This benefit would be even more apparent in the micro-gravity of space, where keeping a paper document in place can be a difficult task. The day of testing did not affect subjects' ability to rectify problems.

[PI] assistance was not significantly helpful for detecting faults. However, Day was found to have a significant main effect on  $T_D$ . That subjects' ability to determine a faulty signal was improved by training implies that training would also improve astronauts' ability to detect instrumentation problems.

Subjects did well in correctly identifying which components of the instrumentation were faulty, while [PI] was generally less correct than subjects in detecting anomalies. The task of correctly detecting faults was sometimes simple enough for subjects to complete without the assistance of [PI]. However, the detection task was made more difficult for Group 2 with [PI] since they already had experience detecting faults without help. The indicator lights were not completely reliable: occasionally a false positive (a red indicator light displayed when the waveform was good quality) or false negative (a green indicator light displayed when the waveform was poor quality) would occur. As a result, subjects were forced to interpret the information from the indicator lights as well as from the waveform output, and this required a greater mental workload than when the indicator lights were turned off. Most subjects preferred the online messages, rather than the NASA guideline, as an aid in diagnosing faults.

Several subjects suggested [PI] could be more helpful if it used a "higher logic" of conjunctive rules between channels and then displayed only those steps that were the *most* likely cause of the problem. For instance, a particular trial involved a reference electrode being removed, which caused several faulty signals at the same time. Even though the diagnosis was *one* faulty component, *three* faulty symptoms were shown by the indicator lights and diagnostic messages. Most subjects were smarter than [PI] on this trial. They skipped the five or more diagnostics steps for a single channel fault and directly asked the assistant to check the reference electrode. The only conjunctive rule now in the [PI] knowledge base manifests when all signals on the display are poor quality.

## CONCLUSIONS

The [PI] software had a beneficial effect in helping subjects troubleshoot faults in sleep instrumentation. Subjects with minimal training preferred having a decision aid to not having one, even if their fault management performance proved otherwise.

[PI] afforded the most benefit with its troubleshooting database. Subjects could quickly look up their diagnostics steps without looking away from the screen. [PI] appears to be an appropriate tool for augmenting conventional training and experiment operation for long-duration missions on the International Space Station.

#### SUMMARY

The expert system, *Principal Investigator-in-a-box*, or [PI], was designed to assist astronauts or other operators in performing experiments outside their expertise. Currently, the software helps astronauts calibrate instruments for a Sleep and Respiration Experiment without contact with the investigator on the ground. It flew on the Space Shuttle missions STS-90 and STS-95. [PI] displays electrophysiological signals in real time, alerts astronauts via the indicator lights when a poor signal quality is detected, and advises astronauts how to restore good signal quality. Thirty subjects received training on the sleep instrumentation and the [PI] interface. A beneficial effect of [PI] and training reduced troubleshooting time. [PI] benefited subjects on the most difficult scenarios, even though its lights were not 100% accurate. Further, questionnaires showed that most subjects preferred monitoring waveforms with [PI] assistance rather than monitoring waveforms alone. This study addresses problems of complex troubleshooting and the extended time between training and execution that is common to many human operator situations on earth such as in power plant operation, and marine exploration.

*Acknowledgements.* - This research was supported by the National Space Biomedical Research Institute, NASA Cooperative Agreement NCC 9-58, and the NASA Ames Research Center, grant number NCC 2-570. The writers would also like to acknowledge the Harvard Medical School Brigham and Women's staff, especially Dr. Charles Czeisler, Dr. Kenneth Wright, and Eymard Riel, for providing training and guidance with the sleep instrumentation.

#### REFERENCES

1. ATAMER, A. *Reliability Evaluation of an Expert System Diagnostic Aid for a Sleep and Respiration Experiment. Master of Science.* Massachusetts Institute of Technology, February 2001.
2. CALLINI, G., ESSIG, S.M., HEHER, D., AND YOUNG, L.R. Effectiveness of an expert system for astronaut assistance on a sleep experiment. *Aviat. Space and Environ Med.*, **71**: 1-10, 2000.
3. DELANEY, M. *Ground-Based Study of an Expert System for Human Assistance on the STS-95 Sleep and Respiration Experiment. Master of Science.* Massachusetts Institute of Technology, February 2001.
4. FRANIER, R.J., GROLEAU, N., HAZELTON, L.R., ET AL. PI-in-a-Box: a knowledge-based system for space science experimentation. *AI Magazine*, **15**: 39-5, 1994.
5. JONES, P.M., AND MITCHELL, C.M. Human-computer cooperative problem solving: theory, design, and evaluation of an intelligent associate system. *IEEE Transactions on*

*Systems, Man and Cybernetics*, **25**: 1039-1053, 1995.

6. ROTH, E.M., BENNETT, K., AND WOODS, D.D. Human interactions with an 'intelligent' machine. *International Journal of Man-Machine Studies*, **27**, 479-525, 1987.
7. ROUSE, W.B. Human problem solving performance in a fault diagnosis task. *IEEE Transactions on Systems, Man, and Cybernetics*, **8**: 258-271, 1978.
8. YOUNG, L.R. PI-in-a-Box. *Journal of the Society of Instrument and Control Engineers*, **33**: 119-22, 1994.