American Nuclear Society Conference 'Computer-based Human Support Systems: Technology Methods and Future'
Philadelphia, Pennsylvania: 1995 June 25-29

ANNUNCIATION IMPROVEMENTS - ASSESSMENT APPROACHES AND LESSONS LEARNED

M.P. Feher and E.C. Davey
AECL, Chalk River Laboratories,
Chalk River, Ontario, Canada, K0J 1J0.
(613) 584-3311

ABSTRACT

AECL, in partnership with Canadian CANDU utilities, has implemented a program to improve the design of annunciation systems. Several innovations were developed, prototyped, and evaluated resulting in considerable interest from several CANDU stations. The need to assess and demonstrate the operational potential of these annunciation improvements has led to the use of several approaches to evaluating and assessing usability and performance. Also, the scarcity of subjects and facilities for testing has led to the development of cost-effective approaches.

This paper discusses approaches used for both formal and informal assessments of the usability and effectiveness of the proposed annunciation system improvements. The paper also discusses the lessons learned from applying the different approaches. Three types of evaluation approaches are presented:

- subject observation - using audio-video recordings to create both the simulation and context of an event to assess the users performance when observing a re-creation (once removed) rather than being immersed in the actual task,
- partial subject interaction - a partial interface prototype in conjunction with audio-video scenario replay to provide context (dynamic non-interactive part-task),
- part-task - a recreation, in real-time, of a part of the system (dynamic interactive part-task), and
- full interaction (immersion) - a modified full-scope simulator to include various configurations of the new concepts (full dynamic interactive).

Finally, the paper presents an approach to a phased implementation of these proposed changes to annunciation.

INTRODUCTION

AECL, in partnership with Canadian CANDU utilities, has implemented a program to improve the design of annunciation systems[1]. Several innovations were developed, prototyped, and evaluated resulting in considerable interest from several CANDU stations. The need to assess and demonstrate the operational potential of these annunciation improvements has led to the use of several approaches to evaluating and assessing usability and performance. Also, the scarcity of subjects and facilities for testing has led to the development of cost-effective approaches.

The development lifecycle of any product should include continuous evaluation and redevelopment with increasing fidelity and formality of the methods used as the product becomes more complete. This is done to ensure that the design is rapidly focused to maximum effectiveness as early in the design process as feasible, and to quantify potential benefits as early in the design process as feasible. To this end, AECL has employed several evaluation techniques during the development of the improved CANDU Annunciation Message List System. The evaluation techniques employed include:

- once-removed,
- dynamic non-interactive part-task,
- dynamic interactive part-task, and
- full dynamic interactive testing.

The overall objectives of the annunciation improvement validation program are to:

- perform validation and evaluation trials of various elements/concepts of the new annunciation strategy,
- incorporate feedback from the validation and evaluation trials to make further improvements in CANDU alarm annunciation,
- investigate and provide recommendations on the integration of the various annunciation facilities into existing station environments,
- establish benefits and risks of a specific configuration prior to implementation of design changes, and
- reduce the regulatory risk of a retrofit to existing stations.

The evaluation techniques are discussed in subsequent sections and a summary of merits and lessons learned is included at the end of the paper.

To better understand the context of the improvements in CANDU annunciation, a brief description of the CANDU Annunciation Message List System (CAMLS) follows.

Summary of CAMLS

AECL has developed an improved computerized annunciation system for the control rooms of nuclear generating stations[2]. CAMLS will alert operators to changes in plant conditions that may impact on safety and production and help staff to effectively respond. CAMLS is designed to:

- provide a clear and concise overview of the current problems or faults in the plant,
- provide an overview of the current state of the plant in terms of automatic process and equipment actions,
- provide support for specific operational tasks, through either pre-configured or operator-configured annunciation displays, including
  - rapid and efficient upset response
  - plant stabilization
  - problem diagnosis
  - recovery action planning and implementation
  - rapid recovery from trip and return to power operation.

To achieve this, CAMLS:

- prioritizes relevant alarm data according to the consequence to the plant and the urgency for an operator response,
- adjusts the alarm presentation and priority with variations in the operating state of the plant,
- significantly reduces irrelevant alarm messages without losing key information,
- improves operator accuracy and speed of diagnosis and planning by providing organized information, and
- prevents distraction from important operational activities through less intrusive and demanding operator interactions.

Currently, CAMLS is a <u>prototype</u> developed for the purpose of validation and evaluation trials of the operational effectiveness of the underlying CAMLS concepts.

CAMLS has two distinct components–two central overview displays and a desktop inquiry system (annunciation interrogation workstation, AIW).

ONCE-REMOVED EVALUATION

As part of the project, known limitations of current annunciation were documented, and the current use was characterized in order to provide a method of comparing the existing designs with current industry-wide standards and to establish a performance benchmark[3]. This work provided a strong basis for improvement to current CANDU annunciation systems. A process of system evaluation and feedback to the R&D process was an integral part of this effort. A once-removed approach to characterizing the situation awareness aspects of existing annunciation systems was developed as a cost-effective approach to overcoming availability and cost constraints (e.g., simulator, trainer, and operator availability) and for collecting operational assessments of annunciation system performance.

Two reference event scenarios (Regional Overpower and Loss of Feedwater) were run on the Point Lepreau Generating Station (PLGS) training simulator. An operating crew participated as actors in a videotaped simulation of the response to the scenarios. A script (for the actors and the simulator program) for these scenarios was developed from a task analysis of operator upset response. The task analysis was used to identify pertinent annunciation and operator actions, and to plan camera angles necessary to capture this information. A simulation of the alarm message list displays was also created directly from simulator data obtained during the scenario taping. The approach used to determine situational awareness was to identify and imbed into the videotaped scenarios annunciation messages that were important enough that an operator should be aware of their occurrence. Sixteen messages were identified by the operator team members as alarms that were feasible yet unrelated to the trip itself. Good operating practice requires all of these alarms to be noted or responded to. If the appropriate action is not performed in time, the consequences of the inaction could pose a threat to plant systems and equipment. Where these alarms appeared in a scenario, the operator/actor in the video would either acknowledge receipt of the alarm or ignore the message and therefore the alarmed condition.

Seven subjects agreed to participate in the study. Two different levels of operators were used. Senior Power Plant Operators (SPPOs) are the licensed authorities for directly controlling the plant, and Power Plant Operators (PPOs) are authorized to monitor control room panels but only perform actions under the direct supervision of the SPPO. Operator-participants who viewed the regional over-power (ROP) trip scenario saw 359 alarm messages on the computer screens, 135 of which were unique strings. The participants were run one at a time, and the sessions were videotaped. The subjects were tested for their situational awareness (i.e., ability to identify important alarms independent of the main upset) during pertinent parts of the scenario. When an operator identified a failure of the operator/actor in the video to respond to an alarm correctly, the tape and annunciation stream were stopped, and the operator’s comments were logged. Several random stops were included, and the operators were queried.

The results indicate that the current annunciation system design results in operators having difficulty in identifying and tracking secondary events/conditions that are not in concert with the flow of the primary problem being addressed. This was demonstrated by the fact that subjects were able to identify less than half of the important independent alarms in the video (note that they were instructed about the nature of the video taped operator errors and were told that their primary task was to find them). SPPOs were able to identify a little more than half of the errors, while the PPOs found less than one quarter, on average. These results, despite the small number of subjects, are certainly worthy of consideration in improved alarm annunciation designs.

This approach has been described as an example of how to gather important information through an evaluation process that can be inexpensive and without requiring extensive simulator or operations staff time. Based on the results with this technique, it should be possible to use this approach in many ways to gather early performance related data for complex systems without substantial need for simulators and the associated training and maintainers necessary to run high fidelity tests.

DYNAMIC NON-INTERACTIVE

Another form of testing that requires minimal simulation and support for testing is non-interactive part-task testing. This form of testing focuses on a small part of the functionality of a larger system to achieve detailed local testing. This type of testing is essential to the design of complex systems to ensure that the individual building blocks of a system perform to the degree required. Note, this type of testing does not address the integration of multiple elements of functionality.

As part of the annunciation improvement program, tests of recognition of abbreviations and acronyms and understanding of message strings was performed. In addition, the program included the development and assessment of prototypes that embodied proposed improvements to the system design. The tests performed during these activities are examples of dynamic non-interactive testing. Each example is briefly discussed below.

Recognition of Abbreviations and Acronyms

Two tasks were used to assess the operators' recognition of abbreviations and acronyms used in alarm message strings. In the first task, the computer displayed an abbreviation or acronym on a screen that was otherwise devoid of any of the other information that comprised a typical message string. In the second task the entire alarm string was presented, with the abbreviation or acronym that was to be identified encased in a box.

As expected, for abbreviations and acronyms displayed out of context, operators were able to interpret fewer terms than when they were provided within the usual message string. The first task, which shows the terms alone, devoid of any other context, indicates that these terms are difficult to interpret (slightly more than half correct on average). Operators performed better in the second task, where the entire message string was presented (three-quarters correct on average).

In both tasks, subjects responded more slowly when they made an error than for correct responses, and SPPOs correctly identified more terms than the PPOs. Since SPPOs are more experienced than their PPO team members, it makes sense that the SPPOs would out-perform them. Normally, the alarm management is performed by the SPPO. However, for upset management, an important part of the alarm screening role is performed by the PPO. Thus, accurate and timely support for alarm interpretation is essential for both SPPOs and PPOs.

These numbers do not represent an assessment of a good or bad design. Rather, they are to be used as a benchmark for comparison with future designs, and as an indication of an opportunity for improvement in abbreviation and acronym use.

Understanding Message Strings

Two tasks were designed to elicit more detailed information about the ability of the message strings to transfer important information to the operators.

In the first task, subjects were presented an alarm string in the middle of the computer screen, after which it disappeared. Subjects were instructed to study the message while it was visible, and then to describe the message in as much detail as possible after it was gone. Elements of each message were identified and required responses defined by operators attached to our project team. The defined elements were assigned one point each for correct interpretation by the test subjects. A test set consisted of 23 different messages and was worth a total of approximately 100 points. Individual messages ranged from a low of three elements to a high of eight elements. The results of this test indicate that operators were able to comprehend two-thirds of the elements of the messages when they were presented alone.

The second task was the most difficult task in the testing series. It was designed to simulate the problems that arise from using scrolling computer screens to display messages during periods of heavy annunciation traffic. This task consisted of a set of seven messages posted to the computer screen one at a time, at two-second intervals. The messages for this task were chosen as cohesive sets of messages that could occur together in an upset scenario. After the last message was posted, the entire set would remain on the screen for an additional seven seconds. At that point, all seven messages were removed from the screen, and operators were instructed to describe the alarm state of the 'plant' by identifying the same message elements and their meaning as in the previous task.

When all subjects were considered, the message design resulted in operators being able to identify and comprehend about one-fifth of the information presented to them. In light of these results, it is not hard to understand why, during an upset, operators are forced to abandon the scrolling annunciation message lists. In this case, they are dependent on annunciation window tiles and alarm printouts produced by the control computer for annunciation information. As in all of the previous message string clarity tasks, the SPPOs got more elements correct than the PPOs.

These results support the statements operators have made about the ineffectiveness of the current message list portion of the annunciation system during upsets. As previously noted, operators had much more time to study the alarms on the screen than they would in a real plant upset, where hundreds of messages come in and cause the message list screens to scroll. Overall, the results of the alarm string clarity testing indicate an opportunity for improvement of the alarm string text and will be used to assess future design improvements.

Assessing Improvements to the Design

A prototyping approach was used for the rapid demonstration and evaluation of new annunciation concepts. Concepts to prototype were selected from the on-going annunciation development work. Annunciation prototypes were developed using Unix-based workstations, and interface builder and plant simulation software. Development and evaluation of each prototype was limited to less than three months to facilitate exploring several options.

Three types of activities were used to obtain comments from participants on the new annunciation concepts:

- Discussions involved single or groups of participants examining a specific concept proposed by the annunciation team. Most of the time this involved review and verbal comment on notes about a specific concept. In several cases, specific exercises were developed to structure the participants comments. For example, participants were asked to group and sort proposed consequence factors in relation to operations goals to determine the content and ordering of factors for the prioritization approach.

- Demonstrations involved groups of participants observing a pre-recorded upset scenario. Annunciation message list displays were simulated by computer displays and a television monitor provided an overview of control room displays and crew communication. Relevant comments from participants were recorded both during the scenario playback and from the discussion afterwards. A typical demonstration and discussion session lasted 40 to 60 minutes in duration.

- Evaluations involved a single participant observing a pre-recorded upset scenario as previously described. In addition to recording relevant comments, participants were asked questions concerning plant and fault state recognition, upset relevant tasks, and how features of the displays and information organization would support them in performing relevant upset tasks.

Overall, the feedback obtained on the concepts was primarily subjective based on each participants operations or design experience.

DYNAMIC INTERACTIVE PART-TASK

Dynamic interactive part-task evaluations involve the evaluation of elements of a design through the use of:

- prototypes or mock-ups of either a whole or part of a system,

- real-time simulation of the parts of a system related to the tasks to be evaluated, and

- interaction with the prototype.

One of the elements of the CAMLS annunciation system, the annunciation interrogation workstation (AIW), was evaluated for its effectiveness, using this approach. The experiment was carried out at Point Lepreau GS station. This section summarizes the approach taken, the analysis of the data collected and the interpretation of the results.

Experimental Procedure

Operators were told to assume they are on shift in the control room and that the plant will experience an upset. They were asked to perform a number of tasks associated with upset diagnosis and response recovery. The tasks performed represented a mix of tasks that could be performed by the senior power plant operator, assistant power plant operator, or shift supervisor. Finally, the operators were told that for some tasks they would be asked to use the normal control room resources. For other tasks they would be asked to use the Annunciation Interrogation Workstation. The subjects were split into two groups. The schedule of activities for Group 1 was as follows:

<u>Group 1</u>

- explanation of experimental procedures and supporting staff roles (5 minutes),

- completion of operational experience questionnaire (5 minutes),

- evaluation of tasks using <u>control room resources</u> (20 minutes),

    - provide scenario starting context to subject

    - begin scenario

- six tasks interspersed throughout the scenario including finding:
    - specific alarm response procedures, cause of trip, conditioning for an alarm, cause of the upset, specific alarm setpoints, and confirming safety system actions,

- questionnaire to gather information on relative support of control room paper-based resources for each task (10 minutes),

- training in AIW functions and practice (10 minutes),

- evaluation of tasks <u>using AIW</u> functions (20 minutes),
    - provide scenario starting context to subject
    - begin scenario
    - seven tasks interspersed throughout the scenario including finding:
        - specific alarm response procedures, cause of trip, conditioning for an alarm, cause of the upset, specific alarm setpoints, confirming safety system actions, and history of alarms,

- questionnaire to gather information on (10 minutes),
    - relative support of AIW resources for each task
    - a comparative assessment of the support provided for each task by the AIW versus current control room resources
    - other tasks AIW could support, changes to current functionality, or additional functions it should include.

<u>Group 2</u>

Group 2 subjects used the same basic experimental procedure except that the AIW was used first and the current control room resources were used second and training was adjusted accordingly.

Summary of Findings from Experiments

Based on the results the Annunciation Interrogation Workstation (AIW) has demonstrated to a high degree of confidence that, compared to the existing support for annunciation related tasks, the AIW improves operator performance by:

- directly supporting tasks for which there was no previous explicit support, and

- clearly has the potential to better support diagnosis, procedural, and information search tasks including:
    - trip cause identification
    - upset cause identification
    - confirmation of automatic responses
    - confirmation of successful safety system trip including trip inhibits
    - access to alarm response procedures
    - access to alarm conditioning, setpoint, and other related information.

FULL DYNAMIC INTERACTIVE TESTING

Full dynamic interactive evaluations involve the evaluation of a full system or sub-system design through the use of:

- prototypes or mock-ups of the system or sub-system,

- ‘real-time’ simulation of the system performance as related to the tasks to be evaluated, and

- subject interaction with the prototype.

One of the elements of the CAMLS annunciation system, the two central annunciation displays, was evaluated using this approach.

The experiment was carried out at the Point Lepreau Generating Station full scope simulator. The focus of this experiment was on validating the effectiveness of the central annunciation displays compared to the existing functionality. This section summarizes the approach taken, the analysis of the data collected and the interpretation of the results.

This validation effort compared the existing station annunciation system with the complete CAMLS system concept across two scenarios. Note, later efforts will begin to isolate the individual CAMLS concepts.

Experimental Procedure

Operators were told at the beginning of each session, that they were to perform the role of SPPO. Also, that they will be supported by other control room operational staff, including a Shift Supervisor (SS). Subjects were drawn from the control room shift complement, refresher training programs, personnel in-training for licensed positions, and from licensed station staff not on shift. The supporting roles were played by a member of the training staff and members of the validation team.

During each scenario, subjects were asked by the SS a series of questions about the plant's state, problems, state trend, safety concerns and production concerns. The interaction between subjects and SS was designed so as to be consistent with normal operational practices. The questions were scenario specific and the answers were recorded in the checklists. Answers provided that were not in the checklist were noted, but were not included in the data analyzed. For the Loss of Feedwater scenario, the time taken for the diagnosis of the root cause of the upset was recorded. After each scenario, the subjects completed a series of anchored rating scales. The subjects were asked to check anywhere along the scale and to use the behavioural descriptor as a guide. After each session, the subject was asked to fill a second questionnaire that provide direct comparative assessment of the support provided by CAMLS or the current central annunciation system. Half the subjects did scenario 1 first and half did scenario 2 first. Each entire session for each subject took approximately 80 minutes. The subjects were split into two groups. The schedule of activities for Group 1 was as follows:

<u>Group 1</u>

- explanation of the experimental procedure and the supporting staff roles (10 minutes),
- completion of questionnaire on operational experience (5 minutes),
- an experimental trial using the existing annunciation system with one scenario with simultaneous collection of objective performance measures (20 minutes),
- completion of subject questionnaires (subjective measures) using anchored rating scales (5 minutes),
- a second trial with the existing annunciation system with the other scenario with simultaneous collection of objective performance measures (20 minutes),
- completion of subject questionnaires (subjective measures) using anchored rating scales (5 minutes),
- explanation of the CAMLS system using the Regional Over Power (ROP) scenario as an example (5 minutes),
- replay of the second scenario with the CAMLS system (5 minutes), and
- completion of the comparative performance questionnaire by each subject (5 minutes).

<u>Group 2</u>

- explanation of the experimental procedure and the supporting staff roles (10 minutes),
- completion of questionnaire on operational experience (5. minutes),
- explanation of the CAMLS system using the ROP scenario as an example (5 minutes),
- an experimental trial using the CAMLS system with one scenario with simultaneous collection of objective performance measures (20 minutes),
- completion of subject questionnaires (subjective measures) using anchored rating scales (5 minutes),
- a second trial with the CAMLS system with the other scenario with simultaneous collection of objective performance measures (20 minutes),
- completion of subject questionnaires (subjective measures) using anchored rating scales (5 minutes),
- replay of the second scenario with the existing annunciation system (5 minutes), and
- completion of the comparative performance questionnaire by each subject (5 minutes).

Summary of Findings from Experiments

Based on the results from the data analysis, it can be said with confidence that CAMLS central annunciation has demonstrated that compared to the existing central message lists plus the window tiles, CAMLS improves:

- the probability of detecting significant alarms,
- the probability of detecting significant problems,
- the probability of detecting alarms secondary or independent of the primary or initial upset,
- awareness of plant state,
- the performance of operators by reducing the demand on operators' short term memory and the resulting mental workload,
- awareness of automatic actions in the plant, and
- the availability of operators for important activities by reducing distracting and unnecessary interaction with the annunciation system.

The first three points noted above have direct safety and economic implications. They can be said to point to an increase in the margins to safety of a CANDU plant and a decrease in the probability of plant trips and equipment damage thereby resulting in an economic saving. The last four points indicate an improvement in human performance in the system. At this point the link between improved human performance in these areas and improved safety and economics is tenuous within the context of this evaluation. However, research in the international aviation industry clearly points to a strong link between these types of measures and the eventual measures of safety and cost effectiveness[4].

SUMMARY OF MERITS OF ASSESSMENT METHODS

Each of the approaches used has practical benefits and pose limitations on the level of confidence in the results obtained and the cost of the evaluation.

The once-removed approach can have significant start-up costs that can only truly be recovered if a large number of subjects are used and when most of the material (e.g., video, audio, simulations) is reusable. If this is true, then a large number of tests can be performed without continued access to simulators nor the need to produce partial simulation.

Dynamic non-interactive testing is a powerful method for early testing of concepts and approaches. This technique can test many of the underlying issues of the design with very simple simulation. It is imperative that designers use the information collected in this phase as a guide and not an absolute. The use of a system with and without interaction is significant when human cognition is an issue. Once again, this is a powerful and relatively inexpensive method of tailoring a system's design to better meet the needs of the intended users.

Dynamic interactive part-task testing can be used as the vehicle for fully understanding the quality of support for specific and isolated tasks. For maximum benefit, the tasks must be clearly defined and understood and the relationship between these tasks other interdependent tasks must be appreciated. If done effectively, this form of evaluation can significantly reduce the scope and challenge of final integrated systems validation.

Full dynamic interactive testing is clearly the method of choice for integrated systems testing. The understanding of the issues of cognitive interactions between sub-systems is clearly best done in this dynamic high fidelity interactive environment. It is equally clear that this forum is not the most cost-effective for measuring many of the intermediate risks and benefits of a system design. Detailed assessment of cognitive issues associated with sub-system designs is often best performed using the other assessment approaches.

LESSONS LEARNED AND CONCLUSIONS

It has been concluded that a continuous process of evaluation with increasing levels of formality throughout the design process is essential to the cost-effective improvement of the design of systems.

It was also clear that to properly assess the human performance related issues of a system requires trained and experienced persons in human factors, plant operation, and process design. As part of any validation experiment, it is necessary to identify performance hypotheses to be tested. The identification of hypotheses related to human performance can not be achieved by following a procedure or using a checklist. The selection of good measures of the hypotheses and scenarios required to establish the data necessary to collect the measures and validate the hypotheses requires that the human factors/evaluation specialists working closely with plant design and operations specialists. This is the only way meaningful results can be obtained.

Early validation need not focus on all measures of system quality. Concept level validation can be done without full consideration for ergonomics (sizes, colours, etc.) of the design. In this way, time can be saved for the more detailed ergonomic design features later in the process when the concepts are more representative of the product to be designed.

Early evaluations during the development lifecycle actually contribute to the design documentation itself. Hypotheses generated for testing the early system prototypes often become the functional and performance requirements for the final product engineering phase.

AECL confirms the value of continuous lifecycle evaluation and will continue to explore new and innovated ways to do cost-effective evaluations throughout this lifecycle.

REFERENCES

1. Feher, M.P., Davey, E.C., and Lupton, L.R. (1994). A User-oriented Approach to Improving Alarm Annunciation. *1994 American Nuclear Society Summer Conference*. New Orleans, Louisiana.

2. Davey, E.C., Feher, M.P., and Guo, K.Q. (1995). An Improved Annunciation Strategy For Candu Plants. To be presented at *1995 American Nuclear Society Summer Conference and Embedded Topical Meeting 'Computer-based Human Support Systems: Technology, Methods and Future'*. Philadelphia, Pennsylvania.

3. Sheehy, E.J., Davey, E.C., Fiegel, T.T. and Guo, K.Q. (1993). Usability benchmark for CANDU annunciation - lessons learned. *1993 American Nuclear Society Topical Meeting on Nuclear Plant Instrumentation, Control and Man-Machine Interface Technologies.* Oak Ridge, Tennessee.

4. Gilson, D.G., Garland, D.J., and Koonce, J.M. (eds) (1994). Situational Awareness in Complex Systems. *Proceeding of the Centre for Applied Human Factors in Aviation (CAHFA) conferenc on Situational Awareness in Complex Systems, Orlando, Florida, February 1-3.*

ACKNOWLEDGMENTS

We would like to thank the management and staff of the Point Lepreau Generating Station (PLGS) in New Brunswick, Canada, for access to and strong input from their operators, trainers, system maintainers and engineers and the full-scope training simulator.